Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services? h) What are the effects of absence or abandonment of grazing on moorland biodiversity and other ecosystem services?

Study details	Authors	Keiller, S.W., Buse, A. & Cherrett, J.M.
	Year	1995
	Aim of study	
	Study design	2
	Quality score	+
	External validity	+
Population and setting	Source population	Some broad descriptions of upland vegetation types and sheep grazing rates in the introduction
	Eligible population	Nine long-established study areas throughout Snowdonia represent a range of upland vegetation types from heathy *Nardus-Festuca* to herb rich *Agrostis-Festuca*. Two of these chosen for this study. Further existing plots in a grazing experiment in Mid-Wales also used.
	Inclusion and exclusion	The study utilises pre-existing exclosures to ensure long-term ungrazed treatments, so

	criteria	choice of study area was limited and dictated by past decisions.
	Setting	Llyn Llydaw, 426m asl Snowdonia, and Pwllpeiran experimental farm, nr Aberystwyth, mid-Wales. Paddocks at around 600m asl.
Methods of allocation to intervention/control	Methods of allocation	Four replicates of treatments in Snowdonia. Randomization not indicated, but appears unlikely. Methods of original selection not described. New grazed treatment areas added, at 40m from the plots. Mid-Wales site part of an existing grazing study. No replication.
	Intervention description	Long-term grazing exclusion (35 years and 16 years, the latter with and without summer grazing in preceding 19 years). Mid-Wales site grazed at ESA stocking rates (given) for different periods, and an ungrazed treatment. In place only for 3-4 years.
	Control/comparison description	Control at Llyn Llydaw is the open hill grazing level and regime. At Pwllpeiran the light grazing treatment could be considered as a control as similar to ESA/ conservation grazing regime. The heavy grazing regime is concentrated in an untypically short period.
	Sample sizes	Invertebrate sampling over 1 year. Surface active arthropods from pitfall traps. Only two per replicate (8 per treatment) at N Wales sites, and 15 per treatment in mid-Wales in three randomly-located transect groups. Diptera emergence from traps at mid Wales (5 per treatment) and Snowdonia podzolic site (4 per treatment). Trap contents recovered every two weeks.
	Baseline comparisons	Baseline pre-dates experiment at Snowdonia site. The paddocks were different at the start – this was an important aspect of the study. History of the mid-Wales paddock is less clear. Dominant species recorded in August 1993 indicate differences in paddocks during study, with ungrazed and lightly grazed sites having more tall heath species and more heavily grazed plots grass dominated and generally shorter.
	Study sufficiently	No power analysis reported.

	powered	
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Numbers and classifications of carabid beetles and spiders, and flies (diptera)
	Secondary outcome measures	Diptera biomass
	Follow-up periods	At mid-Wales site treatments only in place for 3-4 years. At Snowdonia, ungrazed treatments have been in place for 16+ years so long enough to detect long-term effects. Sampling carried out over one year.
	Methods of analysis	Rank-abundance plots for carabids and spiders from different treatments. Shannon – Weiner diversity indices calculated. TWINSPAN analysis of carabid and spider groups and CCA on abundance data. ANOVA of effects on Soil pH, Veg height and dominance etc. on Diptera emergence.
Results		The number of carabid species did not differ markedly between treatments or sites at the Snowdonia or mid-Wales. However number of individuals was highest in the grazed treatments. This may be due to increased probability of trapping carabids in the shorter swards. Spider abundance was generally higher in the ungrazed treatments. At the podzolic site number of spider species was higher in the ungrazed treatments, but higher in the grazed treatments at the brown earth site and in mid-Wales. At the podzolic site Shannon-Weiner diversity for carabids was highest in the grazed treatment, but the 35 year ungrazed treatment for spiders. At the brown earth site spider diversity is highest at the grazed site, but lowest for carabids in this treatment. In mid-Wales, Shannon-Weiner diversity Index was greater in the ungrazed treatment for carabids, and in the ungrazed and lightly grazed treatments for spiders, with no significant difference between these two.

		Un grazed treatments on podzolic soils tended to be dominated by heather and bilberry with grazed treatment s mat grass and sheep's fescue. On the brown earths tall grasses and perennial herbs dominated the ungrazed treatments. Ph was lower in the ungrazed treatments at the north Wales sites. Withdrawal or reduction of grazing pressure will lead to an increase in the diversity and abundance of spiders on podzolic soils, where the vegetation is likely to change from dominance of grasses to vegetation increasingly dominated by heathland species, offering a wide range of available niches. Reduction or withdrawal of grazing will also favour those carabid and spider species associated with heathland and those typical of undisturbed densely vegetated habitats. Diversity is often viewed as a positive conservation feature, but in the species-poor uplands may indicate colonisation by invasive species associated with degraded habitats. It is important to characterise the communities. Differences were more marked in Snowdonia sites, which had been ungrazed for longer, including between the long-term ungrazed podzol and brown-earth plots. In mid-Wales, there were less clear differences between the two grazed treatments. Whilst there is some heather regeneration in the light-grazed treatment, structure and diversity had not developed to the extent where it resulted in heathland arthropod communities. The ungrazed treatment has some characteristic heathland species after three years. There are obvious and expected seasonal variations in dipteran emergence, peaking in late summer. *Nematoceran* biomass in Snowdonia was significantly greater in the grazed treatment, and abundance was higher, but non-significant. This influences overall biomass and abundance with is higher in the grazed treatment. Treatment/ date interactions are significant for most measures. At the mid-Wales site abundance and biomass of most groups were significantly higher in the ungrazed plot, but there was no significant treatment/ date interaction. This increased productivity may be down to increased soil moisture and decaying organic matter.

		The lack of difference in the Snowdonial plots may be down to the high grazing pressure (previously estimated at 2.9 ewes $^{ha-1}$) and resultant dung residue. Sheep dung in heavily grazed areas may act to increase dipteral productivity. Preliminary results of the effect of grazing and dung on subterranean arthropod meso-fauna suggested that withdrawal of grazing might reduce abundance. In summary, the results suggest that a reduction in sheep grazing has little effect on species-richness of carabids and spiders, although species composition will change, in favour of those typical of heather dominated, shaded and densely vegetated habitats. Calculations of diversity, which takes account of both measures, suggests that it increases on podzolic soils with removal of grazing. Spider numbers are likely to increase with a decrease in sheep grazing. Beetles of the family *Scarabaeidae*, obligate dung feeders, are likely to decrease with sheep reductions. Although less clear, the tendency for Diptera is to increase in abundance and biomass with grazing removal.
Notes	Limitations identified by author	Pitfall trapping as essentially a passive method, affected by density of vegetation. Small size of plots and Llyn Llydaw, and possible edge effects.
	Limitations identified by review team	Lack of replication at Pwllpeiran site, limited vegetation measurements of composition and structure.
	Evidence gaps and/pr recommendations for further research	More work needed to understand arthropod preferences for different habitat types. Better understanding of grazing levels to maintain or regenerate upland heathland to benefit characteristic arthropod fauna. The first year of further work on dung effects on soil invertebrates is reported, but not covered above due to limited results at this stage.
	Sources of funding	CCW

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: _____Upland_____

Name of Review Sub-topic (if any): ____Moorland grazing_____

Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services? h) What are the effects of absence or abandonment of grazing on moorland biodiversity and other ecosystem services?
Study Citation	Keiller, S.W., Buse, A. & Cherrett, J.M. (1995). Effects of sheep grazing on upland arthropods in Snowdonia and mid-Wales. CCW Contract Science Report No 120.
Study Design Category	2
Assessed by & when	D Martin 28/11/12

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐+	Comments: Some broad descriptions of upland vegetation types and sheep grazing rates in the introduction
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Comments: Nine long-established study areas throughout Snowdonia represent a range of upland vegetation types from heathy *Nardus-Festuca* to herb rich *Agrostis-Festuca*. Two of these chosen for this study. Further existing plots in a grazing experiment in Mid-Wales also used.
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐+	Comments: Representative of a sub-set of upland vegetation – *Nardus-Festuca* on podzolic soils and *Agrostis-Festuca* on brown earths. The ungrazed paddocks on podzolic soils had developed a heathy vegetation that differed from the surrounding grazed vegetation. A further study area on established paddocks on peaty podzols in mid-Wales. The study utilises pre-existing exclosures to ensure long-term ungrazed treatments, so choice of study area was limited and dictated by past decisions.

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐-	Comments: Four replicates of treatments in Snowdonia. Randomization not indicated, but appears unlikely. Methods of original selection not described. New grazed treatment areas added, at 40m from the plots. Mid-Wales site part of an existing grazing study. No replication.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: Long-term grazing exclusion (35 years and 16 years, the latter with and without summer grazing in preceding 19 years). Mid-Wales site grazed at ESA stocking rates (given) for different periods, and an ungrazed treatment. In place only for 3-4 years.
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐+	Comments: Different periods of exposure at the two sites.
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐++	Comments: Not reported
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐++	Comments:
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Comments:
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: Comparison at Snowdonia site was open hill grazing – grazing level not stated but will reflect prevailing agricultural grazing, possibly modified by agr-environment schemes. At mid-Wales the summer-grazed Apr-Oct treatment most likely to reflect typical

		ESA agreement or conservation grazing regime. The heavy grazing regime is concentrated in an untypically short period.

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐+	Comments: Invertebrate sampling over 1 year. Surface active arthropods from pitfall traps. Only two per replicate (8 per treatment) at N Wales sites, and 15 per treatment in mid-Wales in three randomly-located transect groups. Pitfall trapping has been criticised as a means of estimating population densities. Diptera emergence from traps at mid Wales (5 per treatment) and Snowdonia podzolic site (4 per treatment). Carabid and spider species identified to species level, and other arthropod groups to varying levels. Dipterans identified to sub0order and body-length class.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++	Comments: Trap contents recovered every two weeks.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments:
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐++	Comments:
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐+	Comments: Similar within sites, but different time since stock removed between sites.
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Comments: At mid-Wales site treatments only in place for 3-4 years. At Snowdonia, ungrazed treatments have been in place for 16+ years so long enough to detect long-term effects.

Section 4: Analyses		
4.1 Were exposure and comparison groups		Comments: Baseline pre-dates experiment at

similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐+	Snowdonia site. The paddocks were different at the start – this was an important aspect of the study. History of the mid-Wales paddock is less clear. Dominant species recorded in August 1993 indicate differences in paddocks during study, with ungrazed and lightly grazed sites having more tall heath species and more heavily grazed plots grass dominated and generally shorter.
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐NR	Comments:
4.3 Were the estimates of effect size given or calculable?	☐NR	Comments:
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐++	Comments: Rank-abundance plots for carabids and spiders from different treatments. Shannon – Weiner diversity indices calculated. TWINSPAN analysis of carabid and spider groups and CCA on abundance data. ANOVA of effects on Soil pH, Veg height and dominance etc. on Diptera emergence.
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	Comments: p values given for differences in diversity indices and for dipteran biomass.
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐+	Comments: Replication at the two N Wales blocks, but not at the mid-Wales site.
5.2 Are the findings generalisable to the wider source population(s)/area(s) and	☐+	Comments:

nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?		

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	Do hill sheep occupy a home range and possess a group structure within that range?

Study Details	Authors:	Lawrence, AB and Wood-Gush, DGM.
	Year:	1998
	Aim of study:	To test the home-range behaviour and social organisation of Scottish Blackface sheep.
	Study design:	2
	Quality Score	+
	External validity:	++
Population and setting	Source population:	UK upland hill flock. Not described in detail
	Eligible Population:	62 sheep, part of traditionally farmed flock of pure Scottish blackface sheep. Older ewes culled in September and replaced by ewe lambs. Rest of lambs removed in August. No breakdown of numbers of different age groups within population. Not representative of source population as only one breed of sheep used. Lack of

		information on age groups make it impossible to compare with source population.
	Inclusion & exclusion criteria:	
	Setting:	Area of Pentland Hills (approx. 110 hectares), Nr Edinburgh, Scotland. Reaching approx. 450m a.s.l. Vegetation typical of UK upland heath.
Methods of allocation to intervention / control	Methods of allocation:	62 sheep from flock grazing study area identified with unique individual marks.
	Intervention description:	3 morning or 3 afternoon scan samples (human observation taking two hours per scan, mornings between 06.00 and 12.00 and afternoons between 12.00 and 18.00) performed on seperate days throughout the scan season. During the scan the the location within study area of the marked and unmarked sheep was recorded. Scan seasons Summer 1981(21th May -19th Aug.) , Autumn 1981(13th Oct – 17th Nov), Winter 1982 (2nd Feb-5th April) and Summer 1982 (4th May-23rd Aug.). Recording of unmarked sheep made in one specified area.
	Control / comparison description:	No control.
	Sample sizes:	One 110 ha site with location of 62 marked sheep and some unmarked sheep recorded three times daily (morning or afternoon) over four scanning seasons during a 15 month period.
	Baseline comparisons:	Home range size, Spatial relationships between marked and unmarked sheep, seasonal variation in the distribution of marked sheep, variability in ranging behaviour and consistency of home range behaviour of marked sheep.

	Study sufficiently powered	No power analysis given.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Primary outcome measures:	Home range behaviour and social organisation of sample group.
	Secondary outcome measures:	
	Follow-up periods:	Assessed over total of 12 months, no follow up period.
	Methods of analysis:	Sub group membership determined by fixed distance of 30m. Convex polygon method used to estimate home range size. Cluster analysis used to analyse individuals home range behaviour. Seasonal home range data was put into matrixes using Euclidean distance as the measure of dissimilarity which were then analysed using the Clustan version of Wards Error Sum of Squares method of cluster analysis. Standard deviation was used to compute Inter cluster variability. The one sample X^2 test was used to compare expected values with the observed for cluster compositional stability between seasons.
Results		Results suggested a strong ability of the marked and unmarked sheep to discriminate between each other. There was large seasonal variation in the distribution of the marked sheep with the grazing area increasing during the summer period. Home ranges were also significantly smaller in the winter period. During the winter period clusters were strongly segregated between age classes with those with the majority of the ewe lambs and gimmers ranging more extensively. Younger sheep were markedly less consistent in their home range behaviour between seasons. There was a consistency of membership between clusters in the summer periods of 1981 and 1982 but some sheep were found to move between clusters.
Notes	Limitations identified by	Effect on range behaviour of feed block use in winter period discounted due to return of sheep to winter range area before blocks put out.

	author:	
	Limitations identified by review team:	Short study period, only 1 seasonal replicate and one habitat type may make it difficult to extrapolate results over UK population. Return to winter range area by sheep may be learned response to use of feed blocks in previous years.
	Evidence gaps and/pr recommendations for further research:	Longer study period/ study on other upland habitats. Use of different sheep breeds.
	Sources of funding:	Department of Agriculture and Fisheries for Scotland.

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	Do hill sheep occupy a home range and possess a group structure within that range?

Study Details	Authors:	Lawrence, AB and Wood-Gush, DGM.
	Year:	1998
	Aim of study:	To test the home-range behaviour and social organisation of Scottish Blackface sheep.
	Study design:	2
	Quality Score	++
	External validity:	++
Population and setting	Source population:	UK upland hill flock. Not described in detail
	Eligible Population:	62 sheep, part of traditionally farmed flock of pure Scottish blackface sheep. Older ewes culled in September and replaced by ewe lambs. Rest of lambs removed in August. No breakdown of numbers of different age groups within population. Not representative of source population as only one breed of

		sheep used. Lack of information on age groups make it impossible to compare with source population.
	Inclusion & exclusion criteria:	
	Setting:	Area of Pentland Hills (approx. 110 hectares), Nr Edinburgh, Scotland. Reaching approx. 450m a.s.l. Vegetation typical of UK upland heath.
Methods of allocation to intervention / control	Methods of allocation:	62 sheep from flock grazing study area identified with unique individual marks.
	Intervention description:	3 morning and 3 afternoon (performed of separate days) scan samples of marked sheep noting their location within study area. Scan seasons Summer 1981(21th May -19th Aug.) , Autumn 1981(13th Oct – 17th Nov), Winter 1982 (2nd Feb-5th April) and Summer 1982 (4th May-23rd Aug.). Recording of unmarked sheep made in one specified area.
	Control / comparison description:	No control.
	Sample sizes:	One 110 ha site with location of 62 marked sheep and some unmarked sheep recorded three times daily over four scanning periods during a 15 month period.
	Baseline comparisons:	Home range size, Spatial relationships between marked and unmarked sheep, seasonal variation in the distribution of marked sheep, variability in ranging behaviour and consistency of home range behaviour of marked sheep.

	Study sufficiently powered	No power analysis given.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Primary outcome measures:	Home range behaviour and social organisation of sample group.
	Secondary outcome measures:	
	Follow-up periods:	Assessed over total of 12 months, no follow up period.
	Methods of analysis:	Sub group membership determined by fixed distance of 30m. Convex polygon method used to estimate home range size. Cluster analysis used to analyse individuals home range behaviour. Seasonal home range data was put into matrixes using Euclidean distance as the measure of dissimilarity which were then analysed using the Clustan version of Wards Error Sum of Squares method of cluster analysis. Standard deviation was used to compute Inter cluster variability. The one sample X^2 test was used to compare expected values with the observed for cluster compositional stability between seasons.
Results		Results suggested a strong ability of the marked and unmarked sheep to discriminate between each other. There was large seasonal variation in the distribution of the marked sheep with the grazing area increasing during the summer period. Home ranges were also significantly smaller in the winter period. During the winter period clusters were strongly segregated between age classes with those with the majority of the ewe lambs and gimmers ranging more extensively. Younger sheep were markedly less consistent in their home range behaviour between seasons. There was a consistency of membership between clusters in the summer periods of 1981 and 1982 but some sheep were found to move between clusters.

Notes	Limitations identified by author:	Effect on range behaviour of feed block use in winter period discounted due to return of sheep to winter range area before blocks put out.
	Limitations identified by review team:	Short study period, only 1 seasonal replicate and one habitat type may make it difficult to extrapolate results over UK population. Return to winter range area by sheep may be learned response to use of feed blocks in previous years.
	Evidence gaps and/pr recommendations for further research:	Longer study period/ study on other upland habitats. Use of different sheep breeds.
	Sources of funding:	Department of Agriculture and Fisheries for Scotland.

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Grazing
Review Question	a. Effect of grazing on delivery of moorland biodiversity

Study details	Authors	Littlewood et al.
	Year	2006a
	Aim of study	To determine the effectiveness of moorland vegetation restoration for aiding the restoration of associated insect populations
	Study design	Quantitative observational 2
	Quality score	+
	External validity	+
Population and setting	Source population	Upland heathland
	Eligible population	Not recorded
	Inclusion and exclusion criteria	Sites with moorland restoration through grazing exclusion or herbicide/reseeding
	Setting	8 sites Peak District to Perthshire (grid refs given)
Methods of allocation	Methods of allocation	N/A

to intervention/control	Intervention description	Vegetation restoration at 4 sites through grazing exclusion and 4 sites through herbicide application and reseeding
	Control/comparison description	Comparison between grazing exclusion and herbicide application/reseeded sites
	Sample sizes	18 vegetation sampling positions at each site, hemiptera sampled twice at each site, lepidoptera sampled at a total of 120 trapping events
	Baseline comparisons	N/A
	Study sufficiently powered	+
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Insect counts – treatment method, vegetation condition (percentage cover vascular plants and bryophytes)
	Secondary outcome measures	soil bulk density, pH, loss on ignition, % moisture content, altitude, latitude, precipitation, temperature
	Follow-up periods	N/A
	Methods of analysis	canonical correspondence analysis
Results		Gradient of hemiptera data from degraded to target samples showed separation between treatments (p=0.02) Spp distribution different between nardus and molinia dominated swards Sites with the three highest hemiptera restoration success ranks were all mechanically managed with less patchy regrowth of calluna

		After variation in treatment method, biogeographic variables were more significant than soil variables in explaining hemiptera data Calluna at 2x2m quadrat scale was the single variable than explained most variation in hemiptera assemblage Four of the five highest ranked sites for restoration success of lepidoptera were managed by grazing exclusion. Sample compositions were also affected by geographic location, but no plant variables were significant in explaining variation in assemblages.
Notes	Limitations identified by author	Lack of association with soil variables may have been due to multi-colinearity with treatment as other studies in N. Britain have shown an influence of soil conditions Light traps for lepidoptera sample relative moth activity rather than abundance
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	Possible further studies on influence of vegetation height/ influence on microclimate (humidity etc) on assemblages
	Sources of funding	NERC and Macaulay Institute

Quality Assessment Checklist: Quantitative Study Observational / Correlation v2.0

Name of Evidence Review: _____UPLAND_____

Name of Review Sub-topic (if any): _____GRAZING_____

Review Question	a. Effect of grazing on delivery of moorland biodiversity
Study Citation	Littlewood et al. (2006a)
Study Design Category	Quantitative observational 2
Assessed by & when	SUSANNA PHILLIPS 25/10/2012

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Upland heathland/grassland - N. Stricta/Molinia or calluna Peak district to Perthshire
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: 8 sites – grid reference given and dominant initial species
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: 18 sampling positions at each field site 6 in degraded moorland, 6 in restored calluna moorland and 6 in 'target' calluna moorland 2x2m quadrats Method of selection of soil samples not described

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: 4 sites restoration by grazing exclusion, 4 sites restoration by herbicide & re-seeding
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Hemiptera & Lepidoptera assemblage Vegetation composition, soil variables, altitude, latitude, precipitation and temperature
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐++ ☐+ ☐- ☐NR ☐NA	Comments:
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Counts expressed as a proportion of total sample and zero counts ignored to reduce variability caused by weather conditions or flight season.
2.5 Is the setting applicable to the UK?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Sites in N. England/Scotland

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Visual estimate of percentage cover of species in 2x2m quadrat (and 81 recordings of dominant species at 10m radius of each sample point) Hemiptera sampled twice at each sample point (suction) Lepidoptera sampled by light trapping, total of 120 trapping events
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐++ ☐+ ☐- ☐NR ☐NA	Comments:
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Vegetation compositions assessed, but not vegetation heights
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Direct measures
3.5 Were there similar follow up times in exposure and comparison groups?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Comparison between habitat group (degraded, restored, target), rather than temporal study

3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: See above

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Only 6 samples per site per vegetation condition category (ie degraded/restored/target) but replicated across 6 sites Hemiptera sampled twice at each site
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Insect counts – percentage cover vascular plants and bryophytes, soil bulk density, pH, loss on ignition, % moisture content, altitude, latitude, precipitation, temperature
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Data analysed by canonical correspondence analysis Plant categories present in fewer than 5 samples/sum of ground cover less than 0.8m^2 removed to reduce risk of spurious correlation with rare species. Lepidoptera data included variable for number of days since start of data collection to account for variation in sampling dates between sites
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: P-values given

Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐++ ☐+ ☐-	Comments:
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐++ ☐+ ☐-	Comments:

Evidence Table

Name of Evidence Review:	Upland
Name of Review Sub-topic (if any):	Grazing
Review Question	a. effect of grazing on delivery of moorland biodiversity and other ecosystem services h. effects of abandonment

Study details	Authors	Littlewood et al
	Year	2012
	Aim of study	To consider the impact of different grazing levels on auchenorrhyncha assemblage
	Study design	Quantitative experimental 1
	Quality score	++
	External validity	++
Population and setting	Source population	Upland semi-natural acid grassland, mire M23, M25, U4, U5 and U20
	Eligible population	Floristic diversity as described above
	Inclusion and exclusion criteria	Not reported
	Setting	Glen finglas, Scotland

Methods of allocation to intervention/control	Methods of allocation	Replicated randomised experiment
	Intervention description	Four grazing treatments applied – described in sufficient detail to replicate
	Control/comparison description	Ungrazed plots as control
	Sample sizes	Sample size: 24 enclosures (4 treatments arranged in 6 replicate blocks) 5 randomly selected locations for invertebrate sampling per treatment, leading to 110 D-vac and 79 sweep net samples (total of 3319 adult auchenorrhyncha species)
	Baseline comparisons	No baseline survey reported
	Study sufficiently powered	Not reported
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Invertebrate abundance, species richness and assemblage
	Secondary outcome measures	N/A
	Follow-up periods	Grazing treatments applied from January 2003 & sampling carried out 1 Jun-9 Jul 2007
	Methods of analysis	Generalised Linear Mixed Models to test for significance of grazing treatment on abundance and on species richness Proportion of variation n species data that could be explained by grazing treatment assessed using a Redundancy Analysis (RDA)

Results		D-vac samples showed grazing treatment had a highly significant effect on abundance and a significant effect on species richness (highest median abundance in ungrazed treatment and lowest in high-intensity sheep grazed). Grazing treatment had significant effect on species assemblage data (p=0.001), and the model explained 42.6% of variation. Sweep-net samples showed the grazing treatment effect on species richness and abundance was not significant. Grazing treatment was not significant on species assemblage
Notes	Limitations identified by author	Results represent data collected in single season
	Limitations identified by review team	Results represent data collected from single study site Pre-experiment grazing levels/management may impact on samples, but not reported
	Evidence gaps and/pr recommendations for further research	Replicate study temporally & spatially
	Sources of funding	Scottish Government Rural and Environment Research and Analysis Directorate

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: _____UPLAND_____

Name of Review Sub-topic (if any): _____GRAZING_____

Review Question	a. effect of grazing on delivery of moorland biodiversity and other ecosystem services h. effects of abandonment
Study Citation	Littlewood et al (2012)
Study Design Category	Quantitative experimental 1
Assessed by & when	SUSANNA PHILLIPS 08/11/2012

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Upland semi-natural acid grassland, mire M23, M25, U4, U5 and U20 Vegetation condition not described Glen finglas, Scotland, grid reference provided
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Floristic diversity as described above
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: 24 enclosures each measuring 3.3 ha 4 treatments arranged in 6 replicate blocks 5 randomly selected locations for invertebrate sampling

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Randomised block experiment
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Four grazing treatments applied – described in sufficient detail to replicate Comparisons appropriate – altitude and aspect similar within replicate blocks, but varied between blocks
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Exposure assumed to be as described
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: No other management interventions recorded
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Animals removed during severe weather – assumed to be consistent across plots

2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☒++ ☐+ ☐- ☐NR ☐NA	Comments: Uk based study
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☒++ ☐+ ☐- ☐NR ☐NA	Comments: Range of stocking rates similar to those used across UK moorlands, stock removal in severe weather and for dipping

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++ ☒+ ☐- ☐NR ☐NA	Comments: 2 invertebrate sampling methods – D-vac and sweep-netting, samples subsequently identified to species except female Delphacidae, Aphrodinea and Cixius (aggregated as one group) Inter-rater reliability/QA not reported
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++ ☒+ ☐- ☐NR ☐NA	Comments: Poor weather during sampling period prevented 5 locations being sampled in every plot
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++ ☒+ ☐- ☐NR ☐NA	Comments: Appropriate to meet objectives of study

3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☒++ ☐+ ☐- ☐NR ☐NA	Comments: Direct measures of abundance and species richness
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++ ☒+ ☐- ☐NR ☐NA	Comments: Sampling between 1 June and 9 July 2007
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++ ☒+ ☐- ☐NR ☐NA	Comments: 2003-2007

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐++ ☐+ ☐- ☒NR ☐NA	Comments: No baseline survey reported
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐++ ☒+ ☐- ☐NR ☐NA	Comments: Sample size: 24 enclosures (4 treatments arranged in 6 replicate blocks) 5 randomly selected locations for invertebrate sampling per treatment, leading to 110 D-vac and 79 sweep net samples (total of 3319 adult auchenorrhyncha species)

4.3 Were the estimates of effect size given or calculable?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Effect size not reported
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Generalised Linear Mixed Models to test for significance of grazing treatment on abundance and on species richness Proportion of variation n species data that could be explained by grazing treatment assessed using a Redundancy Analysis (RDA)
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: p-vales reported
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐++ ☐+ ☐-	Comments:
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐++ ☐+ ☐-	Comments:

Quality Assessment Checklist: Quantitative Study Observational / Correlation v2.0

Name of Evidence Review: Upland

Name of Review Sub-topic (if any): Moorland Grazing

Review Question	**d**) Over what timescales can grazing-related change in plant structure and diversity be observed or expected? **e**) How is 'under-grazing' defined? What are the effects of low intensity regimes, set to restore small areas of priority habitat within a moorland mosaic, on other parts of the moorland including non-target habitats such as acid grassland? **h**) What are the effects of absence or abandonment of grazing on moorland biodiversity and other ecosystem services?
Study Citation	Littlewood, N. A.; Pakeman, R.J. & Woodin, S.J. (2006) *A field assessment of the success of moorland restoration in the rehabilitation of whole plant assemblages* Applied Vegetation Science 9: 295-306
Study Design Category	2
Assessed by & when	Amy Christie, 15th-24th January 2013

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐+	Comments: Moors dominated by *Calluna vulgaris*: upland British Isles Biodiversity value described in terms of EC Habitats Directive, EC Birds Directive and UK Biodiversity HAP Threats (replacement of *Calluna* by graminoids) to habitat described with ref to Ball et al 1982; Sydes and Miller 1988 and Bardgett et al 1995 Degraded habitat ascribed to afforestation, overgrazing by sheep, atmospheric deposition, neglect, inappropriate burning
1.2 **Is the eligible population or area representative of the source population or area?** e.g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Comments: Reasonable geographical spread within upland areas of UK; no Welsh or SW sites
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐+	Comments: 8 sites selected within 7 moorland areas: Ben Lawers, Bowland, S. Yorks x3, Geltsdale, Cheshire, Rosedale. Reasons for selecting these particular sites not given, so not clear whether any bias here.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐+	Comments: At each site, 18 sample positions established: 6 acid grassland *Molinia* or *Nardus* ('degraded'); 6 restored dwarf shrub ('restored'); 6 long established dwarf shrub ('target') Sample positions represent time sequence from pre-restoration to restored and forward to desired end point of restoration Areas assigned to 3 management categories: degraded; restored and target, were as similar as possible to each other based on a visual assessment of basic hydrology, altitude and aspect All managed as grazing land, and livestock likely to have been major factor in loss of *Calluna* on degraded areas
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Comments: 2x2m quadrat – percentage cover of all vascular plants and moss and lichen species Soil sampled at each sample point; analysed for bulk density; moisture content; organic matter; pH; nitrogen Sample of *Calluna* taken where present within 5m of sample point; analysed for N
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐++	Comments: In all cases restoration carried out on a subsection of the degraded ground; remaining degraded ground resembled the restored plots prior to restoration
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐++	Comments: Latitude, precipitation, temperature noted for each site; altitude noted for each sample
2.5 Is the setting applicable to the UK?	☐++	Comments: Solely UK sites

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measures subjective or objective? How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐++	Comments: Assessing restoration success – how the restored vegetation assemblages relate to degraded and target samples According to Canonical Correspondence Analysis (CCA) degraded samples separated clearly from restored samples; restored samples largely overlapped with target samples. Constraining the CCA by management status (degraded, restored, target) produced a significantly better model than would be expected by chance, indicating that vegetation assemblages were correlated with management status
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐++	Comments: There was a large degree of homogeneity at each site within each management category with most variation being between management categories and between sites
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++	Comments: Degraded and target samples showed distinct differences in vegetation composition. Diversity was greatest in the degraded *N. stricta*-dominated samples with a mean of 15.7 species per quadrat. In restored quadrats species richness averaged 10.7 per quadrat. Restoration of *Calluna* was largely successful. Five sites had restoration success scores of within 10% of the target and all were within 25% Successful restoration of *Calluna* was not necessarily reflected by successful restoration of the wider plant community, which was very variable – from 96.4% to 5.6%
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++	Comments: Difference in restoration success between sites dependent on methodology. Most successful was grazing exclusion Mechanically restored sites achieved at most 39%

		success compared to a minimum of 73% for grazing exclusion sites Difference in moss assemblage depending on restoration method: mechanically restored largely devoid of bryophytes with site means of 1.7 to 6.7% cover compared to 12.3 to 56.7% on grazing exclusion sites
3.5 Were there similar follow up times in exposure and comparison groups?	☐-	Comments: Study fieldwork undertaken summer 2003; start dates of site restoration varied from 1990 to 1997
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐-	Comments: Gap of between approx 13 years and 7 years between this study and dates of site restoration. Perhaps not long enough to assess long-term effects? There is potential for further recovery towards the desired end state over time

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐-	Comments: Sample size is small
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐++	Comments: Vegetation data set, soil, biogeographic and management variables put into model until no unselected data set would significantly improve the performance of the model ($P<0.05$) Nitrogen content analysed based solely on 114 samples for which nitrogen data available
4.3 Were the analytical methods appropriate? Were important differences in follow-up time	☐++	Comments: Used multivariate methods within CANOCO ver. 4.5 Detrended Correspondence Analysis (DCA) used to

and likely confounders adjusted for? Were sub-group analyses pre-specified?		identify most appropriate form of further analysis: found Canonical Correspondence Analysis (CCA) most appropriate All analysis carried out x2: First using full vegetation set; then with *Calluna, Nardus, Molinia* removed – this allows effects of restoration management on remaining vegetation to be established
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++	Comments: In all analyses by CCA scaling focused on sample distances to best represent the relationship of samples to one another; ordination was carried out without detrending. Axis 1 was constrained to represent variation according to management status The success of restoration management at each site was expressed as the % distance that the mean of the restored samples is along axis 1 from the mean of the degraded samples to the mean of the target samples p-values calculated
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐+	Comments: Moorland management strongly determined vegetation assemblages; however, latitude was the single most important variable. Underlying soil conditions significant contribution to variation in vegetation assemblage. Vegetation change may itself influence soil nutrient concentrations.
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐+	Comments: Relatively small sample size and relatively short length of time elapsed since restoration works initiated are reason for + rather than ++ Four of the five most successfully restored sites were those managed by grazing exclusion. Mechanically restored samples contained relatively low cover of bryophyte and had a generally impoverished flora. Could be related to herbicide used in restoration process; may also be influenced by

		effect of nitrogen deposition on bryophytes (all mechanically restored moors located in Peak District). Management by grazing exclusion often produces only a patchy regrowth of *Calluna*, or in some cases, virtually no *Calluna* over the time scale of the projects studied here. Grazing exclusion has been included as a treatment in previous moorland restoration research with no effect on dominance *of Molinia caerulea*. Summer only grazing can reduce extent (Hulme et al 2002). Little measurable effect under the lower stocking rates prescribed by ESA schemes. At very densely M*. caerulea*-dominated sites intensive herbicide and re-seeding techniques may be only option. Demonstrates importance of setting locally relevant targets. Best practice requires knowledge of initial flora and seed bank. Targets for restoration should consider management not only of key species but also that of remaining vegetation. Most complete rehabilitation achieved with cessation of grazing, especially where *Nardus* dominates; *Molinia* may benefit from summer grazing. Where *Calluna* absent and no viable seed bank, or where elevated nutrient levels give graminoids advantage, herbicide and reseeding may be necessary.

Quality Assessment Checklist: Quantitative Study Observational / Correlation v2.0

Name of Evidence Review: ___Upland_____

Name of Review Sub-topic (if any): ___Moorland grazing_____

Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?
Study Citation	Littlewood, NA, Pakeman, RJ & Woodin, SA (2006c). The response of plant assemblages to the loss of Calluna vulgaris from the upland vegetation. Biological Conservation 128, 335 - 345
Study Design Category	2
Assessed by & when	D Martin

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐++	Comments: Upland heather moorland and degraded moorland. Declines in extent in heather summarised and replacement grassland types mentioned. Some generalised description of invertebrate communities. This study focuses on Hempitera as generally sap feeders, hence respond to vegetation change.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Comments: Eligible population is upland heathland and acid grassland on a range of moors in Scotland and N England, Known to have been previously more heather dominated (mix of wet and dry heath sites). NVC communities given in general terms.
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++	Comments: On each moor twelve randomly located sample positions identified, stratified equally by grass and heath dominated. Vegetation sampled in 2x2m quadrat, including cover estimates and sward heights. Hemiptera sampled by suction at 5 positions within vegetation quadrat for 1 min each. One occasion in each of summer and autumn.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐+	Comments: Exposure is to different historic or long-term grazing levels, expressed as surrogate of dwarf shrub cover. Bias minimised by including a range of sites.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++	Comments: Main explanatory variable is vegetation type (heath, dry acid grassland, wet acid grassland). This is a surrogate in part for past or long-term grazing pressure. Soil variables also included – bulk density, LOI, pH and % moisture).
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NR	Comments: Although classified as three broad vegetation types, will in effect be a range of conditions.
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐+	Comments: Natural variation in populations and composition of Hemiptera between sites. Offset by having a number of sites to improve the likelihood of detecting real trends.
2.5 Is the setting applicable to the UK?	☐+	Comments: Yes, although largely Scotland based

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐+	Comments: Measurements made within a short time period (two weeks) across all sites. Vegetation cover appears to be estimated by eye, on a relatively small number of quadrats, and hemiptera sampled in one year only (although two periods).
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐++	Comments: Yes
3.3 Were all important outcomes assessed?	☐+	Comments: Yes – detailed identification of hemiptera species in order to characterise assemblages. Veg

Were all important positive and negative effects assessed?		composition also assessed as an outcome in relation to stated aims. The latter not assessed by a very rigorous method (cover estimates by eye).
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++	Comments: Yes
3.5 Were there similar follow up times in exposure and comparison groups?	☐+	Comments: All sampling on one year so reflecting current conditions. Site conditions will have arisen from different grazing histories.
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐NA	Comments:

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐NR	Comments:
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐++	Comments: Yes, habitat type (which is a surrogate for management including grazing), but also soil variables as identified in 2.2.
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐++	Comments: ANOVA of plant species-richness between habitat types. Ordination techniques (CCA) used to explore further differences in vegetation assemblages, excluding the dominant species. Vegetation type used as main explanatory variables and soil variables as co-variables. Monte Carlo permutation tests used to determine significance of explanatory variables. Differences in Hemipteran samples between grass and heath analysed using similar techniques. Insect count data log transformed to reduce effects of clumped distribution. Stepwise selection and partitioning used to identify significant variables.
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful?	☐++	Comments: Significant differences identified by ANOVA are given p values, as are selection of explanatory variables.

Were confidence intervals and or p-values for the effect estimates given or calculable?		
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐++	Comments: Only one year study and limited botanical survey, but reasonable sample size and rigorous analysis.
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐+	Comments: The main findings that vegetation composition influences hemipteran assemblage and that grass sites had greatest specie richness, are generalisable across UK heath/ grass moors.

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	d) Over what timescales can grazing-related change in plant structure and diversity be observed or expected? e) How is 'under-grazing' defined? What are the effects of low intensity regimes, set to restore small areas of priority habitat within a moorland mosaic, on other parts of the moorland including non-target habitats such as acid grassland? h) What are the effects of absence or abandonment of grazing on moorland biodiversity and other ecosystem services?

Study details	Authors	Littlewood, N. A.; Pakeman, R.J. & Woodin, S.J.
	Year	2006
	Aim of study	A field assessment of the success of moorland restoration in the rehabilitation of whole plant assemblages: how successful is restoration that is focused on a single dominant plant at enabling the reassembly of the whole vegetation assemblage and what factors affect the relative success of such restoration?
	Study design	2
	Quality score	+
	External validity	+

Population and setting	Source population	Moors dominated by *Calluna vulgaris*: reasonable geographical spread within upland areas of UK; no Welsh or SW sites
	Eligible population	8 sites within 7 moorland areas
	Inclusion and exclusion criteria	Reasons for selecting these particular sites not given, so not clear whether any bias
	Setting	Ben Lawers, Bowland, S. Yorks x3, Geltsdale, Cheshire, Rosedale
Methods of allocation to intervention/control	Methods of allocation	Areas assigned to 3 management categories: degraded; restored and target, were as similar as possible to each other based on a visual assessment of basic hydrology, altitude and aspect
	Intervention description	2x2m quadrat – percentage cover of all vascular plants and moss and lichen species Soil sampled at each sample point; analysed for bulk density; moisture content; organic matter; pH; nitrogen Sample of *Calluna* taken where present within 5m of sample point; analysed for nitrogen
	Control/comparison description	NA
	Sample sizes	At each site, 18 sample positions established: 6 acid grassland *Molinia* or *Nardus* ('degraded'); 6 restored dwarf shrub ('restored'); 6 long established dwarf shrub ('target')
	Baseline comparisons	Sample positions represent time sequence from pre-restoration to restored and forward to desired end point of restoration

		Areas assigned to 3 management categories: degraded; restored and target, were as similar as possible to each other based on a visual assessment of basic hydrology, altitude and aspect All managed as grazing land, and livestock likely to have been major factor in loss of *Calluna* on degraded areas
	Study sufficiently powered	NR
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Assessing restoration success – how the restored vegetation assemblages relate to degraded and target samples All analysis carried out x2: First using full vegetation set; then with *Calluna, Nardus, Molinia* removed – this allows effects of restoration management on remaining vegetation to be established Vegetation data set, soil, biogeographic and management variables put into model until no unselected data set would significantly improve the performance of the model ($P<0.05$) Nitrogen content analysed based solely on 114 samples for which nitrogen data available
	Secondary outcome measures	
	Follow-up periods	Study fieldwork undertaken summer 2003; start dates of site restoration varied from 1990 to 1997 Gap of between approx 13 years and 7 years between this study and dates of site restoration. Perhaps not long enough to assess long-term effects?

	Methods of analysis	Used multivariate methods within CANOCO ver. 4.5 Detrended Correspondence Analysis (DCA) used to identify most appropriate form of further analysis: found Canonical Correspondence Analysis (CCA) most appropriate Constraining the CCA by management status (degraded, restored, target) produced a significantly better model than would be expected by chance, indicating that vegetation assemblages were correlated with management status
Results		Moorland management (referred to as degraded; restored and target samples) strongly determined vegetation assemblages; however, latitude was the single most important variable. Underlying soil conditions significant contribution to variation in vegetation assemblage. Vegetation change may itself influence soil nutrient concentrations. Four of the five most successfully restored sites were those managed by grazing exclusion. Mechanically restored samples contained relatively low cover of bryophyte and had a generally impoverished flora. Could be related to herbicide used in restoration process; may also be influenced by effect of nitrogen deposition on bryophytes (all mechanically restored moors located in Peak District). Management by grazing exclusion often produces only a patchy regrowth of *Calluna*, or in some cases, virtually no *Calluna* over the time scale of the projects studied here. Grazing exclusion has been included as a treatment in previous moorland restoration research with no effect on dominance of *Molinia caerulea*. Summer only grazing can reduce extent (Hulme et al 2002). Little measurable effect under the lower stocking rates prescribed by ESA schemes. At very densely *M. caerulea*-dominated sites intensive herbicide and re-seeding techniques may be only option. Demonstrates importance of setting locally relevant targets. Best practice requires

		knowledge of initial flora and seed bank. Targets for restoration should consider management not only of key species but also that of remaining vegetation. Most complete rehabilitation achieved with cessation of grazing, especially where *Nardus* dominates; *Molinia* may benefit from summer grazing. Where *Calluna* absent and no viable seed bank, or where elevated nutrient levels give graminoids advantage, herbicide and reseeding may be necessary.
Notes	Limitations identified by author	None
	Limitations identified by review team	Relatively small sample size and perhaps relatively short length of time elapsed since restoration works initiated (varies from 7 to 13 years)
	Evidence gaps and/pr recommendations for further research	Where *M. caerulea* is dominant, summer only grazing may assist the recovery of suppressed dwarf shrubs thought the *minimum threshold of dwarf shrub cover for recovery to occur under this type of management is unclear*
	Sources of funding	NERC and the Macaulay Institute

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	

Study details	Authors	Littlewood, NA, Pakeman, RJ & Woodin, SA
	Year	2006c
	Aim of study	To what extent is a loss of Calluna associated with a change in the remainder of the vegetation assemblage? To what extent is vegetation change associated with a change in Hemipteran assemblage? Which of the explanatory variable most influence vegetation and hemipteran assemblage?
	Study design	2 Stratified sampling approach
	Quality score	++
	External validity	+
Population and setting	Source population	Upland heather moorland and degraded moorland. Declines in extent in heather summarised and replacement grassland types mentioned. Some generalised description of invertebrate communities. This study focuses on Hempitera as generally sap feeders, hence respond to vegetation change.
	Eligible population	Eligible population is upland heathland and acid grassland on a range of moors in Scotland and N England, Known to have been previously more heather dominated (mix of wet and dry heath sites). NVC communities given in general terms.

	Inclusion and exclusion criteria	Each moor contained areas of dwarf-shrub vegetation with evidence that this had been the case for a number of decades, and nearby areas of grass-dominated degraded heath. On each moor twelve randomly located sample positions identified, stratified equally by grass and heath dominated.
	Setting	Six moors from Moor House in N Pennines to Ben Lawers in Perthshire and Glensaugh in Aberdeenshire.
Methods of allocation to intervention/control	Methods of allocation	Unclear how field locations were identified – probably subjective but with randomised sampling points within sites. Comments: Exposure is to different historic or long-term grazing levels, expressed as surrogate of dwarf shrub cover. Bias minimised by including a range of sites.
	Intervention description	The 'treatment' is vegetation type, which is a response to different grazing histories, although grazing levels are not quantified.
	Control/comparison description	Modified or degraded grass dominated areas are compared to dwarf shrub dominated sites.
	Sample sizes	Vegetation sampled in 2x2m quadrat, including cover estimates and sward heights. Hemiptera sampled by suction at 5 positions within vegetation quadrat for 1 min each. One occasion in each of summer and autumn.
	Baseline comparisons	One-off study.
	Study sufficiently powered	N/A
Outcomes and methods of analysis (inc effect size, CIs for each	Primary outcome measures	Detailed identification of hemiptera species in order to characterise assemblages. Veg composition also assessed as an outcome in relation to stated aims. The latter not assessed by a very rigorous method (cover estimates by eye).

outcome and significance)	Secondary outcome measures	Identification of key environmental variables affecting main outcomes.
	Follow-up periods	One-off assessment, but reflecting long-term effects of different management.
	Methods of analysis	ANOVA of plant species-richness between habitat types. Ordination techniques (CCA) used to explore further differences in vegetation assemblages, excluding the dominant species. Vegetation type used as main explanatory variables and soil variables as co-variables. Monte Carlo permutation tests used to determine significance of explanatory variables. Differences in Hemipteran samples between grass and heath analysed using similar techniques. Insect count data log transformed to reduce effects of clumped distribution. Stepwise selection and partitioning used to identify significant variables.
Results		Plant species-richness was greater in grasslands, and differences in composition with heath samples characterised by a number of mosses. Dry grassland sites were the most species-rich. Grass sites typically had higher soil bulk density and pH, and lower loss on Ignition (LOI). Grass samples also had a greater number of Hemiptera taxa than heathland, and clear differences in the assemblages of the heath and wet and dry acid grassland samples. Nine taxa were most commonly associated with heath, and 25 taxa most closely associated with grass samples. There is however evidence of a greater degree of specialism in the species found in heathland samples. Plant species accounted for more variance in Hemipteran assemblage than other variables, but half of the variation was unexplained. As well as temperature and latitude, loss on ignition and mean vegetation height explained some of the variance. There is broad agreement with other studies that vegetation is the most important factor, with soil explaining further variation. The study suggests that vegetation change and change in associated structure is likely to lead to changes in other aspects of biodiversity, and in this case an increase in Hemipteran diversity. The effects were consistent across a broad range of

		sites with varying latitude, hydrology and altitude.
Notes	Limitations identified by author	Soil factors identified as being key explanatory variables are not necessarily causal factors.
	Limitations identified by review team	One year of study. Effects of grazing are inferred in the vegetation type, but no measurements of grazing pressure or stocking rate.
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	NERC and Macaulay Institute

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?

Study details	Authors	Littlewood, N
	Year	2008
	Aim of study	To investigate the response of nocturnal adult Lepidoptera assemblages to different grazing regimes. Follow on from Dennis et al, 2008.
	Study design	1
	Quality score	++
	External validity	++
Population and setting	Source population	Upland semi-natural grassland. Not described in detail.
	Eligible population	Study area is a large upland area, likely to be typical of a range of grazed upland habitats
	Inclusion and exclusion criteria	
	Setting	

Methods of allocation to intervention/control	Methods of allocation	Six replicates of four treatments. Allocation within blocks randomized. Experimental design is as Denis *et al*, 2008)
	Intervention description	Treatments high sheep -2.7 sheep ha^{-1}, low sheep – 0.9 sheep ha^{-1}, mixed sheep and suckler cows in autumn, to give similar rate to high sheep, and ungrazed control
	Control/comparison description	Ungrazed, or low sheep as continuation of previous management
	Sample sizes	Traps placed at random points within one replicate of each treatment and rotated around, so each treatment on four replicates sampled on 6 occasions over a four month period in the last year of the experiment.
	Baseline comparisons	
	Study sufficiently powered	
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Moth diversity and abundance
	Secondary outcome measures	-
	Follow-up periods	4 month period, but grazing treatments in place for previous 5 years.
	Methods of analysis	Moth counts log transformed and effect of treatment analysed using two-way ANOVA. Date included as a factor to account for differing weather conditions. Similar analysis on species-richness treatment. Moths also split into graminoid and other feeders, and overwintering strategy and BAP spp vs non BAP. Associations between groups and treatments tested by chi-squared testing.

Results		A total of 6291 moths of 153 species of aggregates were recorded. The largest moth abundance was in the light sheep-grazed treatment (52.2 per night) followed by ungrazed (47.9 per night). The heavy sheep grazed treatment had the lowest abundance (33.8 per night) Differences were significant (p=0.029). Species richness was greatest in the ungrazed treatment (13.2 species per night) followed by light sheep-grazed (12.3 species per night) and lowest in the heavy sheep-grazed treatment (10.6 species per night). These differenced were significant (p=0.012). There was a significant interaction between treatment and feeding preferences with more graminoid feeders in the ungrazed plots (p<0.001). There were significantly more species that overwinter as larva or pupae in the ungrazed treatment, and more egg stage in light sheep-grazed (p<0.001). Whilst BAP species were distributed through all treatments, there was a greater than expected proportion in the heavy sheep-grazed treatment ((p=0.025). Overall, the trend was for higher moth species richness in the light sheep grazed and ungrazed plots. Other studies have shown that Lepidoptera are less tolerant of disturbance than are plants. A small number of BAP priority species were however found in the heavily grazed plots. They were however primarily herbaceous species, which may fare better than graminiverous species where heavy grazing suppresses grasses. Graminoid feeders were disproportionately well represented in the ungrazed plots, which were characterised by dense grass tussocks. The lightly grazed plots may provide conditions for species with different preferences. The mixed grazing treatment only differed significantly from the ungrazed. This supports other work at the site which did not show a significant interaction of Lepidopteran abundance with cattle grazing, unlike other invertebrate groups.
Notes	Limitations identified by author	
	Limitations identified by	

	review team	
	Evidence gaps and/pr recommendations for further research	Examining associations between moth species and quantifiable treatment effects within the plots
	Sources of funding	Rural Environment Research and Analysis Directorate of the Scottish Government

DENNIS, P., SKARTVEIT, J., MCCRACKEN, D. I., PAKEMAN, R. J., BEATON, K., KUNAVER, A. & EVANS, D. M. 2008. The effects of livestock grazing on foliar arthropods associated with bird diet in upland grasslands of Scotland. *Journal of Applied Ecology,* 45**,** 279-287.

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: _____Upland _____

Name of Review Sub-topic (if any): _____Moorland grazing_____

Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?
Study Citation	Littlewood, N. (2008) Grazing impacts on moth diversity and abundance on a Scottish upland estate. Insect Conservation and Diversity 1, 151-160
Study Design Category	1
Assessed by & when	D Martin 20/12/12

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐+	Comments: Upland semi-natural grassland. Not described in detail.
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Comments: Study area is a large upland area, likely to be typical of a range of grazed upland habitats
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++	Comments: The vegetation of sample plots is described in terms of dominant NVC types. Three paired replicate blocks of four treatments, each plot 3.3 ha, so fairly representative. Location/ selection not described

DENNIS, P., SKARTVEIT, J., MCCRACKEN, D. I., PAKEMAN, R. J., BEATON, K., KUNAVER, A. & EVANS, D. M. 2008. The effects of livestock grazing on foliar arthropods associated with bird diet in upland grasslands of Scotland. *Journal of Applied Ecology,* 45**,** 279-287.

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐++	Comments: Six replicates of four treatments. Allocation within blocks randomized. Experimental design is as Denis *et al*, 2008)
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: Treatments high sheep -2.7 sheep ha^{-1}, low sheep – 0.9 sheep ha^{-1}, mixed sheep and suckler cows in autumn, to give similar rate to high sheep, and ungrazed control
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐++	Comments: treatments in place for 5 seasons. Reasonable to detect grazing effects.
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐++	Comments:
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐++	Comments:
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Comments: The range of vegetation types in the sample area is widespread in the UK uplands.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++	Comments: Grazing treatments are within the range found in the uplands with the high grazing treatment reflecting the levels that have resulted in problem grazing, and the low rate being more typical of

		sustainable grazing encouraged through schemes.

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++	Comments: Moths sampled using Skinner-type light traps. Traps placed at random points within one replicate of each treatment and rotated around, so each treatment on four replicates sampled on 6 occasions over a four month period in the last year of the experiment.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐+	Comments: Three sampling periods unsuccessful due to equipment failure and poor weather.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐+	Comments: As per objectives, but no vegetation structure or botanical measurements.
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐++	Comments:
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Comments:
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Comments: Five-year study – should be sufficient to detect grazing effect, though may continue to change over a longer timescale.

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐++	Comments: Stated that there were no differences in vegetation communities or structure between treatments at the outset.

4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐NR	Comments:
4.3 Were the estimates of effect size given or calculable?	☐NR	Comments:
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐++	Comments: Moth counts log transformed and effect of treatment analysed using two-way ANOVA. Date included as a factor to account for differing weather conditions. Similar analysis on species-richness treatment. Moths also split into graminoid and other feeders, and overwintering strategy and BAP spp vs non BAP. Associations between groups and treatments tested by chi-squared testing.
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++	Comments: p-values given for all tests
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐++	Comments: Well designed and replicated
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐++	Comments:

Evidence Table

Name of Evidence Review:	Upland
Name of Review Sub-topic (if any):	Grazing
Review Question	a. The effect of grazing on the delivery of moorland biodiversity

Study details	Authors	Marrs et al.
	Year	2004
	Aim of study	To test a range of management treatments to reduce molinia cover and encourage the development of dwarf shrub vegetation
	Study design	Quantitative experimental
	Quality score	+
	External validity	+
Population and setting	Source population	North Peaks/Yorkshire Dales – grid references given. Molinia & Molinia/calluna moorland
	Eligible population	1x white moor (molinia), 1x grey moor (molinia + calluna) in each location 400m^2 block at each site chosen – rationale for selection not reported
	Inclusion and exclusion criteria	Existing grazing regime under ESA/CSS agreement
	Setting	North Peaks & Yorkshire Dales

Methods of allocation to intervention/control	Methods of allocation	2 areas per site, 3 randomly allocated grazing treatments in randomly allocated burnt/unburnt treatments
	Intervention description	Each area (4000m^2) was split into two (2000m^2 each) and randomly burnt or left unburnt. Each 2000m^2 was divided into three grazing plots, and grazed at agri-environment rate, summer only grazing or no grazing. 10x10m plots in each grazing sub-plot were treated with a herbicide treatment (high rate, low rate or no application)
	Control/comparison description	Comparison between sub-polts (as described above)
	Sample sizes	Replicated on 4 moors and replicated on 2 sites per moor. Within each treatment plot, vegetation height measured at 20 sample points, and cover of higher plant species estimated at 4 random 1x1m quadrats. Total of 960 samples of vegetation height
	Baseline comparisons	Exposure to treatments started march 1995, first survey carried out in summer 1995. Data provided suggest similar vegetation characteristics in baseline survey
	Study sufficiently powered	+
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Vegetation height and species composition
	Secondary outcome measures	N/A
	Follow-up periods	1995-2000
	Methods of analysis	Univariate analysis of individual responses and multivariate analysis within framework

		of constrained ordinations for community response
Results		The majority of the study considered effects of burn/herbicide treatments, however, grazing-related results were as follows: the effects of grazing treatments in the study were inconsistent between regions. Grazing had negligible significant effects on sward height over the 6 years, but tallest vegetation was found in the grazing exclusion plots (but low stocking densities and sheep only used in experiment). A greater positive effect was shown on moorland-bog development of grey sites than on white sites
Notes	Limitations identified by author	Low winter stocking rates across site as a whole, grazing levels on plot could not be confirmed and therefore low winter grazing plot may be subject to same treatment as summer-only grazing plot
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	Effect of alternative stocking scenarios
	Sources of funding	Defra funded/Monsanto plc supplied herbicides

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: _____UPLAND_____

Name of Review Sub-topic (if any): _____GRAZING_____

Review Question	a. The effect of grazing on the delivery of moorland biodiversity
Study Citation	Marrs et al (2004)
Study Design Category	Quantitative observational
Assessed by & when	SUSANNA PHILLIPS 01/11/2012

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐++ ☐+ ☐- ☐NR ☐NA	Comments: North Peaks/Yorkshire Dales – grid references given. Molinia & Molinia/calluna moorland No further details of habitat composition or condition
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: 1x white moor (molinia), 1x grey moor (molinia + calluna) in each location 400m^2 block at each site chosen – rationale for selection not explained
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Samples for vegetation height and cover estimates chosen at random - selection of random point was not reported

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: 3 randomly allocated grazing treatments in randomly allocated burnt treatments, 3 areas with herbicide application in each grazing plot – not reported how selected
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Detailed & appropriate experimental design
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Low winter stocking rates across site as a whole, can not confirm grazing levels on plot and therefore low winter grazing plot may be subject to same treatment as summer-only grazing plot
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Not reported, assumed acceptable
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: No other interventions reported

2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Assumed to be typical of molinia dominated moors in UK from limited information provided
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Agri-environment stocking rates typical of moors in England

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Sward stick measurements of vegetation height Calluna seedling count Subjective measure of species % cover – visual estimates – inter-rater reliability not reported
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: All outcome measures described were reported on
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Vegetation heights, covers of key species (incl molinia), calluna seedling emergence – outcomes measured meet objectives of study

3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Direct measures
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Measurements carried out 'in summer' – covers wide timescale, not clear if each plot recorded at same time each year
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: 1995-2000

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Exposure to treatments started march 1995, first survey carried out in summer 1995. Data provided suggest similar vegetation characteristics in baseline survey
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: 4 sites with 2 experimental blocks per site. 3 grazed treatments x 2 burn treatments per site (ie 6 plots per site), 20 samples per plot. Total 960 samples of vegetation height

4.3 Were the estimates of effect size given or calculable?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Effect size not reported
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Individual responses – repeated measures ANOVA Community responses – multivariate analysis – detrended correspondence analysis/canonical correspondence analysis
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: p-values given
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐++ ☐+ ☐-	Comments: Subjective cover estimates Stocking rate for site known but unknown for individual plot
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐++ ☐+ ☐-	Comments: Paper identifies marked variability between apparently similar sites in different regions

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?

Study details	Authors	Martin, D
	Year	2011
	Aim of study	• To assess the condition of these commons near the start of the CSS agreements. This will establish baseline survey information. • To determine what changes have occurred by repeating the surveys in subsequent years of the agreement.
	Study design	2 Repeated random sample survey
	Quality score	+
	External validity	+
Population and setting	Source population	Limited description of the habitat in general, in the English or UK context
	Eligible population	Probably fairly typical of degraded moss heath in English mountains, but greater in area than most stands so may have areas of better condition than most other English stands. Not as extensive as stands in Scotland.
	Inclusion and exclusion	Random sampling approach likely to be fairly representative. However in some years more of

	criteria	the block scree area was included in the sample, which often holds patches of taller moss so may introduce bias.
	Setting	Summit plateau of Cross Fell, North Pennines, Cumbria. Altitude 830-890m.
Methods of allocation to intervention/control	Methods of allocation	Randomly generated sample points within a delineated survey area to include the extent of the target feature.
	Intervention description	Stock reduced over whole moorland area under a grazing agreement, and active regular shepherding of sheep from the summit area, to give an annual average grazing pressure of <0.5 sheep per ha
	Control/comparison description	No direct comparison, other than change form baseline survey.
	Sample sizes	Varied between 20 in 2008 and 44 in 2005. In 2010 55 quadrats were placed but 20+ fell on rocky areas.
	Baseline comparisons	Baseline survey of condition in 2003, just after initial stock reductions. All variables used in the analysis measured at baseline.
	Study sufficiently powered	Power to detect significant change calculated for a number of variables. Variables with large sample size (grass and moss heights etc) high powered. The small sample size in 2008 affected the power of analysis for a number of variables, although graminoid heights okay as four measurements per quadrat increased power.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Frequency of Racomitrium and stiff sedge. Height and cover of moss and stiff sedge. Frequency and height of bilberry. Frequency of lichens.
	Secondary outcome measures	Cover and height of sheep's fescue (improved condition would be an increase of moss and stiff sedge at expense of sheep's fescue).
	Follow-up periods	2003 – 2010. Four surveys over this period.

	Methods of analysis	Analysis of Variance on measures of height and cover of species. Some variables transformed by natural log. Kruskal-Wallis test used where there were a high number of zero measures. Graphs of mean values over time. Bar charts of frequency of each species over time. NVC comparisons.
Results		Over eight years of reduced grazing on Cross Fell plateau, the proportion of sample points dominated by *Carex bigelowii* increased to 52% of quadrats from 14% in 2003, at the expense of *Festuca ovina.* The mean height of all species including *Vaccinium, Carex bigelowii* and mosses increased significantly, with the average graminoid height more than doubling to 6.5cm. Average moss and lichen cover also doubled to 27% from 13%. There was no positive response in lichen species, with declines in some, which may be related to increased competition and shading. The overall trend was for decreasing grass dominance and increased similarity to montane sedge and moss-heath vegetation.
Notes	Limitations identified by author	Differences in sample area boundaries in some years, low sample size in one year, low power of some variables to detect significant change due to small number of observations.
	Limitations identified by review team	Only one site. Grazing pressure not quantified through either observation or surrogate measures.
	Evidence gaps and/or recommendations for further research	Longer monitoring would allow a better assessment of timescales of recovery to favourable condition. Failure of lichens to respond to reduced grazing – how can they be encouraged.
	Sources of funding	NE internal work programme

Quality Assessment Checklist: Quantitative Study Observational / Correlation v2.0

Name of Evidence Review: ___Upland_____

Name of Review Sub-topic (if any): ___Moorland grazing_____

Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?
Study Citation	Martin, D. 2011. *Survey of grazing impact on Cross Fell montane heath, and analysis of change 2003-2010.* Natural England, unpublished Habitats and Plants Team report
Study Design Category	2
Assessed by & when	D Martin

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐+	Comments: Limited description of the habitat in general, in the English or UK context
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Comments: Probably fairly typical of degraded moth heath in English mountains, but greater in area than most stands so may have areas of better condition than most other English stands. Not as extensive as stands in Scotland.
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐+	Comments: Random sampling approach likely to be fairly representative. However in some years more of the block scree area was included in the sample, which often holds patches of taller moss so may introduce bias.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐NA	Comments: time series monitoring of one area, looking at change over time rather than comparing sites or treatments.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Comments: Yes – basic hypothesis is that grazing pressure affects structure and cover of key species in the habitat. Moderate grazing pressure will affect low productivity habitats. However mainly looking at change over time, rather than attempting to correlate with stocking rate data or surrogate measures of grazing.
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NA	Comments:
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐++ ☐+ ☐- ☐NR ☐NA	Comments:
2.5 Is the setting applicable to the UK?	☐++	Comments: Yes, results would be broadly applicable to the habitat throughout the UK, and England particularly.

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐+	Comments: Area was sampled at randomly generated points, aiming for 50, although this was not achieved in some years (min 20). Located using GPS to remove selection bias. Observers assessed some quadrats together at the beginning for consistency. Heights measured accurately using sward sticks, but cover estimates more subjective.
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐+	Comments: All measurements generally completed, although sample size lower in some years.

3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++	Comments: Main elements of condition all measured: over and heights of key species, bare ground, presence of droppings.
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++	Comments: Yes
3.5 Were there similar follow up times in exposure and comparison groups?	☐NA	Comments: Monitoring/ time series study
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐++	Comments: Measured at 4 times from 2003 to 2010, so allows trends to be detected.

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	Comments: Power calculations carried out in MINITAB for different variables. Some variables (heights of dominant species) were sufficiently powered due to large sample size, others were low powered, so less confidence in significant results. Only 20 sample points in 2008 affected the power of analysis for some variables although graminoid height still well powered as four measurements per quadrat increased power. Discussed in a section of the report.
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐-	Comments: Main assumption is that effects are down to change in grazing levels
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐+	Comments: Analysis of Variance on measures of height and cover of species. Some variables transformed by natural log. Kruskal-Wallis test used where there were a high number of zero measures. Graphs of mean values over time. Bar charts of frequency of each species over time. NVC analysis.
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++	Comments: Yes, p-values given for all tests.

Section 5: Summary

5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐+	Comments: Reasonably large randomised sample and survey frequency allows detection of trends. Some issues with differences in sampling area boundary in different years.
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐+	Comments: Findings generally valid across the habitat, in terms of the effects of grazing reduction from previously heavily grazed situation.

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	

Study details	Authors	Medina-Roldán, E., Paz-Ferreiro, J. & Bardgett, R. D
	Year	2012
	Aim of study	To test whether grazing exclusion is associated with a slowing of N and C cycling, characterised by reductions in rates of N mineralisation and microbial activity, and an increase in soil C and N content.
	Study design	2
	Quality score	+
	External validity	++
Population and setting	Source population	Upland grass and heath ecosystems.
	Eligible population	Eligible area is typical of extensively grazed upland acid grassland. The study includes a long-term (8 yrs at time of study) ungrazed area of 170ha – not typical of usual management.
	Inclusion and exclusion criteria	Opportunistic – NNR area with existing grazing excluded area
	Setting	Ingleborough NNR, Yorkshire Dales

Methods of allocation to intervention/control	Methods of allocation	Subjectively chosen study site. Plots were restricted random.
	Intervention description	one area is grazed year round up to 4 ewes ha^{-1} (yr round or summer?) + cattle. Pre-existing regime. Other area has had grazing removed for 8 yrs.
	Control/comparison description	Control is on-going grazed area
	Sample sizes	Grazed padock is 58 ha and ungrazed area is 170 ha. Six plots located at random within a study area, with each plot sub-divided into 16 sub-plots.
	Baseline comparisons	Sampled over one year – not at treatment baseline.
	Study sufficiently powered	N/A
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Above and below ground biomass, size of litter horizon and organic horizons; Soil ammonium and nitrate, total N and C and microbial N and C; DOC; bulk density
	Secondary outcome measures	
	Follow-up periods	All measurements over a year
	Methods of analysis	ANOVA and t-tests on transformed data
Results		Grazing exclusion increased the relative abundance of dwarf shrub and reduced the proportion of graminoids. Dominance of graminoids differed, with Nardus, Festuca and Agrostis in the grazed are and Deschampsia spp and Eriophorum vaginatum in the exclusion area. Above ground biomass was higher in the exclusion area in May and July. Grazing exclusion increased the litter layer by 70%, which may be a result of the greater

		contribution from woody dwarf shrub, and the slower decomposition of this material, and greater accumulation of dead material. Root biomass varied seasonally but there was no grazing effect. Other organic horizons were not affected by grazing or season. Grazing removal caused a 20% reduction in microbial activity and reduced net ammonia mineralisation at all sampling except January of second year. The increased ration of dissolved organic to inorganic N is consistent with a decrease in ecosystem productivity. The slowing down of nutrient cycling may be related to reduced dung inputs, and changes in the character of plant litter. Microbial biomass also decreased by 30% in the ungrazed area. Grazing exclusion was shown not to affect water table depth, so soil moisture was not a significant driver of differences in nutrient cycling. Despite slowing of N cycling and build up of litter, grazing exclusion did not increase C or N in soil. This adds to other evidence of lack of response to grazing removal, even after 30 years (Garnett et al, 2000; Ward et al, 2007) Grazing exclusion did not modify other soil properties including DOC, DON and microbial C, which showed stronger seasonal variation than between grazing treatments.
Notes	Limitations identified by author	Psuedoreplication since only two areas studied
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	Whether restoration of dwarf shrub at the expense of graminoids and non-Sphagnum mosses can increase the potential for C sequestration.
	Sources of funding	Mexican Council for Science and Technology (PhD funding)

Quality Assessment Checklist: Quantitative Study Observational / Correlation v2.0

Name of Evidence Review: ___Upland_____

Name of Review Sub-topic (if any): ___Moorland grazing_____

Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?
Study Citation	Medina-Roldán, E., Paz-Ferreiro, J. & Bardgett, R. D. 2012. Grazing exclusion affects soil and plant commnunities, but has no impact on soil carbon storage in an upland environment. *Agriculture, Ecosystems and Environment,* 149, 118-123.
Study Design Category	2
Assessed by & when	D Martin

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐+	Comments: Upland grass and heath ecosystems.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Eligible area is typical of extensively grazed upland acid grassland. The study includes a long-term (8 yrs at time of study) ungrazed area of 170ha – not typical of usual management.
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++	Comments: Same as eligible area – landscape scale study. Within this area sampling plots were partially randomised within a larger block

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐-	Comments: NNR sites - one area is grazed year round up to 4 ewes ha^{-1} (yr round or summer?) + cattle. Pre-existing regime. Other area has had grazing removed for 8 yrs. Site identification is opportunistic rather than objective.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Comments: Grazing presence or absence. No assessment of the effect of different grazing levels.
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐++	Comments:
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐+	Comments: Sample areas selected to have similar topography and altitude
2.5 Is the setting applicable to the UK?	☐++	Comments: Yes – typical of much extensive upland rough grazing

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Turves to 20cm depth taken from each sub-plot 5 times over just over 1 year, starting in spring. Sampled from 16 sub-plots from each of 6 plots per treatment. Root cores taken from adjacent to sub-plots. Water table measured twice-weekly in each plot. Turves were separated into different above and below ground organic fractions. Above ground biomass was separated into functional groups. Soil bulk density, Total C and N were estimated for two sampling dates. Soil ammonium and nitrate were measured by machine, as was DOC.
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐++	Comments:
3.3 Were all important outcomes assessed?	☐++	Comments:

Were all important positive and negative effects assessed?		
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++	Comments:
3.5 Were there similar follow up times in exposure and comparison groups?	☐++	Comments: All subject to same sampling regime. The grazing/ no grazing regimes will have been in place for different durations
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐+	Comments: Probably – eight years enough to detect differences?

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐NR	Comments:
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐+	Comments: mainly grazing vs grazing exclusion. Sample date also considered as fixed factor, and replicate plots as random factor.
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐+	Comments: ANOVAs and t-tests expressed on area basis using bulk density and soil depth. All variables transformed to normalise them
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++	Comments:
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)?	☐+	Comments: There is only one area of each treatment. Pseudoreplication is mentioned as unavoidable.

How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design		
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐++	Comments: Site is fairly typical of UK uplands.

Quality Assessment Checklist: Quantitative Study Observational / Correlation v2.0

Name of Evidence Review: Natural England Uplands Evidence Review

Name of Review Sub-topic (if any): **What are the effects of grazing regimes and stocking rates on the maintenance and or restoration of moorland biodiversity and ecosystem service delivery**

Review Question	d) Over what timescales can grazing-related change in plant structure and diversity be observed or expected?
Study Citation	Miller, G., Geddes, C. & Mardon, D.K. (1999) Response of the alpine gentian *Gentiana nivalis* to protection from grazing by sheep
Study Design Category	2
Assessed by & when	Tom Holland 6th February 2013

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Yes – a small population of alpine gentians within a small area (one hectare) of calcareous alpine dwarf-shrub CG12 at 1000-1100m on Ben Lawers.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: I guess so – I think the area formed part of the sample on which CG12 description is based.
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: The gentians were so scarce that all clusters containing six or more plants were included within the study (i.e. sixteen quadrats).

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: The gentians were so scarce that all clusters containing six or more plants were included within the study (i.e. sixteen quadrats). Does this mean that there is no selection bias?
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Seemed reasonable to me – I can't think of anything else that you might want to know given the limitations (fundamental limitations?) of the study (i.e.not being able to study the respons of the vegetation to a number of different stocking rates).
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: The gentians were so scarce that all clusters containing six or more plants were included within the study (i.e. sixteen quadrats). Does this mean that there is no contamination?
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: The authors recognise that the measured variables only explain a certain percentage of the variation in gentian numbers. Other variables such as autumn and spring temperatures, amounts of rainfall, strength of wind, duration of snow lie, herbivory by molluscs and rodents are listed as possible confounding factors but were not included within the study.
2.5 Is the setting applicable to the UK?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Asks and attempts to answer a question very specific to alpine gentians in the UK

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: I think so - For example % cover were done with a pin-frame, which seems more accurate that doing it by eye. However,
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: A measurement of bare ground in spring was made for five years (1988-1992) but discontinued when it seemed to become apparent that the autumn measurement had more affect on gentians numbers. Should it have been continued to make sure the lack of correlation real rather than apparent?
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Analysis and discussion of results correlating number of gentians with the number of seed capsules and % bare ground cover the previous September and height of vegetation the previous July are presented in the discussion section and not in the results section.
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: I think so apart from a measurement of bare ground in spring was made for five years (1988-1992) but discontinued when it seemed to become apparent that the autumn measurement had more affect on gentians numbers.
3.5 Were there similar follow up times in exposure and comparison groups?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: I think the ungrazed and grazed plots were surveyed at the same time of year within a week or two.

3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: The study was done over ten years, and seemed to produce plausible results. It would be interesting to know if the study has been continued and whether the general trends continued.

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐++ ☐+ ☐- ☐NR ☐NA	Comments:
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Analysis of covariance, split plot analysis and regression analysis applied to the different parts of the data.
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Seem ok to me.
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: For the main results 95% confidence limits were provided along with the mean. P-value were given for the analysis of results correlating number of gentians with the number of seed capsules and % bare ground cover the previous year.

Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐++ ☐+ ☐-	Comments: Pin frame used to prevent bias in % cover measurements. Weather variables same were not included within the study (for were the same for both plots).
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐++ ☐+ ☐-	Comments: Probably applicable to other alpine gentian populations in the UK and to other plants with a similar way of life (i.e. annuals).

Evidence Table

Name of Evidence Review:		Natural England Uplands Evidence Review
Name of Review Sub-topic (if any):		**What are the effects of grazing regimes and stocking rates on the maintenance and or restoration of moorland biodiversity and ecosystem service delivery**
Review Question		d) Over what timescales can grazing-related change in plant structure and diversity be observed or expected?
Study details	Authors	Miller, G., Geddes, C. & Mardon, D.K. Response of the alpine gentian *Gentiana nivalis* to protection from grazing by sheep
	Year	1999
	Aim of study	To clarify the response of alpine gentians to protection from grazing – more specifically to determine what effect protection from grazing would have on population density, growth, survival and seed production of alpine gentians.
	Study design	Correlation (correlating different grazing regimes with alpine gentian numbers & vegetation structure)
	Quality score	2+
	External validity	EV+
Population and setting	Source population	One hectare of calcareous alpine dwarf-shrub CG12 at 1000-1100m on Ben Lawers
	Eligible population	Eight pairs of 70cm x 50cm quadrats surveyed annually for ten years (1987-1996).
	Inclusion and exclusion criteria	Not known.
	Setting	Calcareous alpine dwarf-shrub CG12
Methods of allocation to	Methods of allocation	Of each pair of quadrats one is open to sheep grazing throughout the year and one is open to grazing in the winter (December to April).

intervention/control	Intervention description	Excluding summer grazing between May to November from half of the quadrats (one of each pair).
	Control/comparison description	Correlating different grazing regimes with changes in vegetation composition and structure.
	Sample sizes	Sixteen 70cm x 50cm quadrats surveyed annually for ten years (1987-1996)
	Baseline comparisons	All quadrats surveyed before grazing excluded from half.
	Study sufficiently powered	Data transformed. Analysis of covariance, split plot analysis and regression analysis applied to the different parts of the data.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Number of gentian flowering and number seeding within each quadrat. Number of flowers and number of capsules on each plant. Average seed content per capsule from a sample taken from outside of the plots. Height of each plant at flowering time. Height of vegetation in plots. Cover of bare ground in plots measured in autumn.
	Secondary outcome measures	
	Follow-up periods	Quadrats surveyed in summer and autumn annually for ten years.
	Methods of analysis	Data transformed. Analysis of covariance, split plot analysis and regression analysis applied to the different

		parts of the data.
Results		For the first three years, numbers of gentians in the ungrazed plots matched those in the grazed plots but after that they declined a lot more than those in the grazed plots. Vegetation height increased in the ungrazed plot (to 50-60cm), but not in the grazed plot, and the cover of bare ground decreased from in the ungrazed plot but not in the grazed plot. These alpine gentians on the ungrazed plots grew taller and survived better (<10% mortality compared to 30-50%) than did plants in adjacent grazed plots. Gentian numbers fluctuated more on the grazed plots than the ungrazed plots. The correlation of gentian number with bare ground cover the previous autumn in the ungrazed plots shows the importance of bare ground for the plant's reproduction. The results suggest that around about 4% cover is needed. The results suggest it should be possible (if not practical on Ben Lawers) to manipulate the stocking rate to maximise the gentian population.
Notes	Limitations identified by author	They note the criticisms made by other researchers of studies comparing simple ungrazed and grazed scenarios. They defend the study by saying that a more sophisticated experiment that controlled stocking levels would be impractical in the location and with such a small area of habitat.
	Limitations identified by review team	Analysis and discussion of results correlating number of gentians with the number of seed capsules and % bare ground cover the previous September and height of vegetation the previous July are presented in the discussion section and not in the results section.
	Evidence gaps and/pr recommendations for	

	further research	
	Sources of funding	

Summary

Miler *et al* (1999) sought to clarify the response of alpine gentians to protection from grazing – more specifically to determine what effect protection from grazing would have on population density, growth, survival and seed production of alpine gentians. Within a hectare of calcareous alpine dwarf-shrub CG12 at 1000-1100m on Ben Lawers, eight pairs of 70cm x 50cm quadrats surveyed annually for ten years (1987-1996).

Miller *et al* (1999) [2+] showed that the numbers of gentians declined more in the ungrazed plots than in the grazed plots over the duration of the ten-year study. Vegetation height increased in the ungrazed plot (to 50-60cm), but not in the grazed plot, and the cover of bare ground decreased from in the ungrazed plot but not in the grazed plot. These alpine gentians on the ungrazed plots grew taller and survived better (<10% mortality compared to 30-50%) than did plants in adjacent grazed plots. Gentian numbers fluctuated more on the grazed plots than the ungrazed plots. The correlation of gentian number with bare ground cover the previous autumn in the ungrazed plots shows the importance of bare ground for the plant's reproduction. The results suggest that around about 4% cover is needed. The results suggest it should be possible (if not practical on Ben Lawers) to manipulate the stocking rate to maximise the gentian population.

Quality Assessment Checklist: Quantitative Study Observational / Correlation v2.0

Name of Evidence Review: Natural England Uplands Evidence Review

Name of Review Sub-topic (if any): What are the effects of grazing regimes and stocking rates on the maintenance and or restoration of moorland biodiversity and ecosystem service delivery

Review Question	d) Over what timescales can grazing-related change in plant structure and diversity be observed or expected?
Study Citation	Miller, G., Geddes, C. & Mardon, D.K. (2010) Effects of excluding sheep from an alpine dwarf-shrub community - Plant Ecology & Diversity Vol 3, No 1, February 2010, 87-93
Study Design Category	2+
Assessed by & when	Tom Holland 6th February 2013

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Yes –a small area (one hectare) of calcareous alpine dwarf-shrub CG12 at 1000-1100m on Ben Lawers.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: I guess so – I think the area formed part of the sample on which CG12 description is based.
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Maybe not, but probably not far off - The location of the quadrats was determined by the location of gentian clusters (on which the main study looking at changing gentian numbers was focussed).

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Selecting the quadrats on the presence of gentian cluster weighs the bias towards whatever factors favour that species.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Seemed reasonable to me – I can't think of anything else that you might want to know given the limitations (fundamental limitations?) of the study (i.e.not being able to study the response of the vegetation to a number of different stocking rates).
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: All quadrats received pretty much the same level of grazing or non-grazing and a suffered similar levels of exposure to the montane climate?
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Plant species grouped into functional groups (dwarf shrubs, graminoids, mat-forming, erect/decumbent forbs, annual forbs, pteridophytes, bryophytes) for analysis.
2.5 Is the setting applicable to the UK?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Asks and attempts to answer a question specific to montane dwarf shrub-grassland community in the UK

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: I think so - For example % cover were done with a pin-frame, which seems more accurate that doing it by eye and difficult plant groups were grouped to avoid mis-identification.
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: A measurement of bare ground in spring was made for five years (1988-1992) but discontinued when it seemed to become apparent that the results were not useful. Should it have been continued to make sure the lack of correlation real rather than apparent?
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Seemed to have been.
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: I think so apart from a measurement of bare ground in spring was made for five years (1988-1992) but discontinued when it seemed to become apparent that the autumn measurement had more affect on gentians numbers.
3.5 Were there similar follow up times in exposure and comparison groups?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: I think the ungrazed and grazed plots were surveyed at the same time of year within a week or two.
3.6 Was the follow up time meaningful?	☐++	Comments:

Was the follow-up long enough to assess long-term effects?	☐+ ☐- ☐NR ☐NA	The study was done over ten years, and seemed to produce plausible results. It would be interesting to know if the study has been continued and whether the general trends continue.

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐++ ☐+ ☐- ☐NR ☐NA	Comments:
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Data analysed by repeated measures analysis of variance.
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Seem ok to me.
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: For the main results 95% confidence limits were provided along with the mean. P-value were given for the analysis of results.

Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐++ ☐+ ☐-	Comments: Pin frame used to prevent bias in % cover measurements. Difficult plant groups were grouped to avoid mis-identification Weather variables same were not included within the study (but were the same for both plots).
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐++ ☐+ ☐-	Comments: Probably applicable to other alpine shrub and grassland populations in the UK.

Evidence Table

Name of Evidence Review:		Natural England Uplands Evidence Review
Name of Review Sub-topic (if any):		**What are the effects of grazing regimes and stocking rates on the maintenance and or restoration of moorland biodiversity and ecosystem service delivery**
Review Question		d) Over what timescales can grazing-related change in plant structure and diversity be observed or expected?
Study details	Authors	Miller, G., Geddes, C. & Mardon, D.K. Effects of excluding sheep from an alpine dwarf-shrub community
	Year	2010
	Aim of study	Determine the role of sheep grazing in conserving the dwarf-shrub community on Ben Lawers, Perthshire, Scotland
	Study design	Correlation (correlating different grazing regimes with changes in vegetation composition and structure)
	Quality score	2+
	External validity	EV+
Population and setting	Source population	One hectare of calcareous alpine dwarf-shrub CG12 at 1000-1100m on Ben Lawers
	Eligible population	Eight pairs of 70cm x 50cm quadrats surveyed annually for ten years (1987-1996).
	Inclusion and exclusion criteria	Not known.
	Setting	Calcareous alpine dwarf-shrub CG12
Methods of	Methods of allocation	The location of the quadrats was determined by the location of gentian clusters (on which the main study looking at changing gentian numbers was focussed). Of each pair of quadrats one is

allocation to intervention/control		open to sheep grazing throughout the year and one is open to grazing in the winter (December to April).
	Intervention description	Excluding summer grazing between May to November from half of the quadrats (one of each pair).
	Control/comparison description	Correlating different grazing regimes with changes in vegetation composition and structure.
	Sample sizes	Sixteen 70cm x 50cm quadrats surveyed annually for ten years (1987-1996)
	Baseline comparisons	All quadrats surveyed before grazing excluded from half.
	Study sufficiently powered	Tests of statistical significance have been carried out on functional groups (dwarf shrubs, graminoids, mat-forming, erect/decumbent forbs, annual forbs, pteridophytes, bryophytes, lichens, litter, bare ground and vegetation height).
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	% cover of plant species Plant species grouped into functional groups (dwarf shrubs, graminoids, mat-forming, erect/decumbent forbs, annual forbs, pteridophytes, bryophytes) for analysis. % cover of bare ground, bryophytes, litter & lichens
	Secondary outcome measures	
	Follow-up periods	Quadrats surveyed annually.
	Methods of analysis	Data analysed by repeated measures analysis of variance.
Results		Excluding sheep caused major shifts in the balance amongst species. The vegetation composition and structure of grazed plots did not change significantly over the course of the experiment, but in the summer-ungrazed plots bryophyte and litter cover

		increased whilst bare ground decreased. The cover of graminoids and some forbs increased in the middle years of the experiment before declining back towards the baseline level by the end of the experiment.
Notes	Limitations identified by author	They note the criticisms made by other researchers of studies comparing simple ungrazed and grazed scenarios. They defend the study by saying that a more sophisticated experiment that controlled stocking levels would be impractical in the location and with such a small area of habitat.
	Limitations identified by review team	Should an examination of changes in NVC community type been done, to see if the summer-ungrazed had become more similar or less similar to the published accounts of CG12? Some CG12 constants such as Alchemilla alpine appear to have done better under the summer-ungrazed regime than the year round-grazed regime.
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	

Summary

Miller et al (2010) sought to determine the role of sheep grazing in conserving the dwarf-shrub heath community on Ben Lawers by comparing areas open to grazing with areas where sheep-proof cages were erected each summer for ten years. Vegetation composition and structure data were collected.

Miller et al (2010) [2+] found that excluding sheep caused major shifts in the balance amongst species. The vegetation composition and structure of grazed plots did not change significantly over the course of the experiment, but in the summer-ungrazed plots bryophyte and litter cover increased whilst bare ground decreased. The cover of graminoids and some forbs increased in the middle years of the experiment before declining back towards the baseline level by the end of the experiment.

Evidence Table

Name of Evidence Review:	Upland
Name of Review Sub-topic (if any):	grazing
Review Question	h. what are the effects of abandonment on biodiversity and other ecosystem services

Study details	Authors	Milligan et al
	Year	2004
	Aim of study	To test the effects of a number of mechanical and chemical treatments on the abundance of Molinia
	Study design	Quantitative experimental 2
	Quality score	+ (- based on grazing element alone)
	External validity	+
Population and setting	Source population	Molinia-dominated moorland
	Eligible population	Molinia habitats at Ramsgill bents, N Yorkshire
	Inclusion and exclusion criteria	N/A
	Setting	N Yorkshire

Methods of allocation to intervention/control	Methods of allocation	Randomised block (n=2) (split-split design)
	Intervention description	2 grazing treatments in each block as main plot treatments (ESA stocking rate plus unknown density of rabbit grazing, and reduced grazing – sheep and rabbit grazing prevented by exclosure fences) 4 sub-plots - cutting treatments in each main plot 2 sub-sub plots in each sub-plot - herbicide or calluna brash treatments
	Control/comparison description	Control – fenced, ungrazed site/no treatments
	Sample sizes	3 quadrats per sub-sub-plot
	Baseline comparisons	Baseline description of species composition – mainly qualitative
	Study sufficiently powered	No power calculation reported Replicated experiment – two main plots, but small sample sizes
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Species cover, bare ground cover, height, litter depth
	Secondary outcome measures	Calluna seedling density recorded from 1997
	Follow-up periods	Cutting carried out from Dec 1995, herbicide and grazing from July 1996 and brash treatments in January 1997. Data collected 1996-1999
	Methods of analysis	ANOVAs to assess affect of treatments on species covers and physiognomic variables

		Multivariate analyses used to describe changes in community composition MANOVA used to determine significance of each explanatory variable
Results		The most successful treatment was cutting x3 and grazing. The effect of cutting 3 times reduced vegetation height, and this effect was still observed 44 months after treatment. Cutting once and 2 times showed similar results to untreated plots after 7 and 19 months. Cutting 3 times significantly reduced litter depth (3.5cm to 0.3cm) Species number and Shannon-weiner index showed significant change in 1999 survey in both cut and grazed plots ($p<0.05$). Both measures greater in ungrazed plots and increased in cut plots with increased intensity of cutting The MANOVA showed that of the main treatment effects, only grazing and time were significant
Notes	Limitations identified by author	Longer term research is needed to assess the most appropriate treatments Effect on fauna unknown, lack of knowledge of effect on invertebrates is particularly noted by the authors
	Limitations identified by review team	Effect of calluna brash treatment likely to be longer term than timescales (except effect on calluna seedling density) studied in this research Study investigated only sheep (and rabbit) grazing treatments The experimental design included replication, but this was carried out on an individual site, more geographically broad data would make conclusions more robust
	Evidence gaps and/pr recommendations for further research	Effect on fauna of each treatment – particularly ground nesting birds/invertebrates with cutting treatments Effect of cattle/pony grazing may also be of interest Cutting treatments only applied to one/two years, effect of successive years cutting

		would be of interest Economic implications of each approach
	Sources of funding	DEFRA

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: _____UPLAND_____

Name of Review Sub-topic (if any): _____GRAZING_____

Review Question	h. what are the effects of abandonment on biodiversity and other ecosystem services
Study Citation	Milligan et al (2004)
Study Design Category	Quantitative experimental 2
Assessed by & when	SUSANNA PHILLIPS 16/11/2012

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Molinia-dominated moorland
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Ramsgill bents, N Yorkshire Grid reference provided
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: 3 permanent 1x1m quadrats in each sub-plot – initial method of selection not reported

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Randomised block (n=2) (split-split design) Each block 50x60m
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: 2 grazing treatments in each block as main plot treatments (ESA stocking rate plus unknown density of rabbit grazing, and reduced grazing – sheep and rabbit grazing prevented by exclosure fences) 4 cutting treatments in each main plot Within these sub-plots, herbicide or calluna brash treatments Field procedures described in detail
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐++ ☐+ ☐- ☐NR ☐NA	Comments: No details given, assumed as experimental design
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: No contamination reported
2.5 Were any other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: No other intervention reported, assumed as per experimental design

2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Sample population may be typical of other molinia-dominated habitats in UK, but the paper notes a large variability in response of molinia communities
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Grazing treatment likely to reflect usual UK practice, cutting may also be carried out on some moor. Herbicide and brash treatments are rare, and generally only where specified by restoration/agri-environment scheme.

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Species cover assessed visually – subjective measure, not reported if validated (quadrat subdivided into 100 sub-sections to increase accuracy) Bare ground cover (subjective), height, litter depth Calluna seedling density recorded from 1997
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: All outcome measures reported on
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++ ☐+ ☐- ☐NR ☐NA	Species cover, bare ground cover, height, litter depth Calluna seedling density recorded from 1997. Study focussed on impact on vegetation, effects on fauna were not included Measures appropriate to address objectives of study

3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐++ ☐+ ☐- ☐NR ☐NA	Direct measures used
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Surveys carried out in same month on each sub-plot
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Experiment ran 1995 to 1999, further effects (particularly of grazing) may have been shown over longer timescales

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Baseline description of species composition – mainly qualitative
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is	☐++ ☐+ ☐- ☐NR	Comments: No power analysis given Replicated experiment – two main plots Small sample sizes – 3 quadrats per sub-sub-plot

the expected effect size? Is the sample size adequate?	☐NA	
4.3 Were the estimates of effect size given or calculable?	☐++ ☐+ ☐- ☐NR ☐NA	Comments:
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: ANOVAs to assess affect of treatments on species covers and physiognomic variables Multivariate analyses used to describe changes in community composition MANOVA used to determine significance of each explanatory variable
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: p-value provided
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐++ ☐+ ☐-	Comments: Replicated design, randomised block (split-split design). Lack of detail on some aspects of study, some degree of subjectivity, and limited sample sizes (ie 3 quadrates per sub-sub plot. Downgraded to '-' based on grazing aspects – only one grazing level – plots open to moorland grazing, plus grazing exclusion. Actual grazing pressure not quatified
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐++ ☐+ ☐-	Comments: Sample population may be typical of other molinia-dominated habitats in UK, but the authors notes a large variability in response of molinia communities

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: _____Upland _____

Name of Review Sub-topic (if any): _____Moorland grazing_____

Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?
Study Citation	Milne, J., Pakeman, R. J., Kirkham, F. W., Jones, I. P. & Hossell, J. 2002. Biomass production of upland vegetation types in England and Wales. Grass and Forage Science, 57, 373-388 Kirkham, F.W. & Milne, J. A. 2000. Progress towards defining ecologically sustainable grazing management: the 'Moorland Biomass' and 'Heather Suppression' projects. Aspects of Applied Biology, 58. Vegetation management in changing landscapes.
Study Design Category	2
Assessed by & when	D Martin 26/1/13

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐++	Comments: A range of upland heath and grass communities extensive in the UK uplands Heather dominated communities
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++	Comments: Covering the main upland areas of England, ITE Land Classification used to select squares at random for the study, three or four per area. Same areas used for heather suppression study
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐+	Comments: Actual 1 ha study plots selected subjectively, for vegetation uniformity, proportion of dominant species, and growth stages of heather. Actual sampling areas selected at random within the blocks Sites chosen to hold a single phase of heather growth and four sites in each chosen to reflect a range of grazing intensity

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐++	Comments: Three sample 3.5m x3.5m enclosures selected at random within the 1 ha plots each year before growth commenced. Different locations in each sample year. Biomass sampled from three randomly positioned quadrats in each exclosure. Likely to have been subjective to cover the range of heather ages and grazing intensities
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++ +	Comments: Basically biomass samples from grazing exclosures. Cover of live vegetation of the target species estimated using intersections on a cross-wired quadrat before cutting. Subjective range of grazing pressures. Broadly replicable. Current season's growth and remaining woody growth sampled each year and adjusted to 100% ground cover. Seasonal exclosures set up in each area
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐+ -/ NR	Comments: Sampled in three seasons, so will take account of some climatic variation.
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐NR	Comments:
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐NR	Comments:
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Comments:
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK	☐NR	Comments:

practice(s)?		

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++	Comments: Biomass sampled to ground level in three random quadrats. In the heather quadrats, non-heather veg was discardred. A 7% sub-sample removed and the remainder oven dried and weighed. The sub-sample was used to get a measure of current seasons shoot as a proportion of overall biomass. Adjusted to give weight per unit area at 100% cover values, using the cover measures. Simolas approach for V myrtillus. Measured in October, and at other points in growing season for one plot in each region. Similar approach for graminoids plots, but sampled at more points in growing season. For Ag-Fe was cut to 4cm three times during the season and cut material measured. Then cut to ground level in Oct. Weight of seasons green shoots and woody portion measured . Grazing Index measured in grazed and ungrazed plots
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐- +	Comments: Not all area sampled in 1996 and 1997 due to resources
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments:
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐++	Comments: Fairly straightforward measures of biomass production.
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐NR	Comments:
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐NR	Comments:

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐NR	
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐NR	Comments:
4.3 Were the estimates of effect size given or calculable?	☐NA	Comments:
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐++	Comments: Unbalanced design with missing values (e.g. not all heather age classes found in each area) conventional analysis of variance not possible. Residual max likelihood used. Ultimately, data for each species analysed separately, with region and year as random effects. Weather variables used to build models of biomass dependence. Simpler random model used grasses, with only one plot per region. C vulgaris analysed with growth phase as a factor. Relationships between GI and growth variables (absolute and ungazed-grazed) tested by linear and polynomial regression
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++	Comments:
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐+ +	Comments: Random selection of study areas and plots, but unbalanced design – not all veg types in all areas

5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐++ ++	Comments: Representative of typical upland communities from most English upland areas

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	

Study details	Authors	Milne, J., Pakeman, R. J., Kirkham, F. W., Jones, I. P. & Hossell, J.
	Year	2002
	Aim of study	To obtain annual and seasonal values for biomass production from a range of vegetation types for England and Wales and to quatify effects of environmental variables on biomass production. To identify the response of heather to variation in grazing intensity, location and development phase.
	Study design	2 stratified random sample, with some replication within sample plots.
	Quality score	+
	External validity	++
Population and setting	Source population	A range of upland heath and grass communities extensive in the UK uplands
	Eligible population	Covering the main upland areas of England, ITE Land Classification used to select squares at random for the study, three or four per area.
	Inclusion and exclusion criteria	Actual 1 ha study plots selected subjectively, for vegetation uniformity, proportion of dominant species, and growth stages of heather. Actual sampling areas selected at random within the

		blocks
	Setting	6 upland areas of England and Wales
Methods of allocation to intervention/control	Methods of allocation	Three sample 3.5m x3.5m enclosures selected at random within the 1 ha plots each year before growth commenced. Different locations in each sample year. Biomass sampled from three randomly positioned quadrats in each exclosure.
	Intervention description	Biomass removal from sample areas in grazing exclosures
	Control/comparison description	N/A
	Sample sizes	For most vegetation types three or four sample areas per region
	Baseline comparisons	Chosen to be uniform and reasonably high cover of target vegetation
	Study sufficiently powered	N/A
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Biomass and cover
	Secondary outcome measures	
	Follow-up periods	Measured over three years
	Methods of analysis	Unbalanced design with missing values (e.g. not all heather age classes found in each area) conventional analysis of variance not possible. Residual max likelihood used. Ultimately, data for each species analysed separately, with region and year as random effects. Weather variables used to build models of biomass dependence. Simpler random model used grasses,

		with only one plot per region. C vulgaris analysed with growth phase as a factor.
Results		There was a significant effect of year on current season's growth for each heather phase. There was also a significant effect of region, with the south west having the lowest biomass of current seasons growth. The older growth phases had significantly greater current growth. No significant interactions of phase with region or year were detected. There was no significant effect of region or year on live annual biomass production of Vaccinium. The amount of live biomass of Nardus in August was significantly influenced by region and year, with the North Pennines having lower values for other regions. There was no significant effect of region or year on Molinia live biomass, but this was not sampled in all regions and in some regions only one square was measured. E vaginatum was sampled in three regions, with the North Pennines having significantly lower values than north east England or the South Pennines. There was no significant effect of region or year on *Agrostis-Festuca* biomass production. There was rapid growth in all phases of Calluna between June and August, with little accumulation between August and October. There was an indication of lower growth between June and August in the south west compared with other regions. The greater biomass production in the north-west region compared with elsewhere is associated with greater growth in the early part of the season, which is then maintained. A range of environmental variables were significant in explaining biomass production, including soil density and total number of frost days in the previous winter for Calluna. Mean monthly maximum daily temperature explained significant variation in the biomass production of Nardus. The mean biomass productivity of different heather growth phased was found to be higher than previous Scottish-based studies (e.g. Grant et al, 1982), which may have implications for the utilization thresholds applied throughout the country, and may affect the carrying capacity of heather before heather growth is suppressed., or the productivity could lead to a rapid increase in the woody portion, making plants more susceptible to grazing. The greater biomass production seems to be from a greater rate

		of production early in the growing season, and possibly a function of the length of growing season. Unlike previous studies a higher productivity was found in mature plants.
Notes	Limitations identified by author	Comparability of results with other studies due to variations in methodology, although methods applied here seem robust. Some question over how senescence affects biomass results (net rather than gross). Frequency of defoliation on Ag-Fe and effect on growth. There was considerable variation in the relationship between GI and both shoot growth and differential productivity. Differential shoot growth of pioneer heather was positively related to GI when data for all regions and years were included, indicating an overall suppression in growth relative to ungrazed heather. Mature heather growth in grazed areas showed a quadratic relationship to GI overall which was significant when both variables averaged over three years. Data for pioneer heather in 1998 suggested an increasing differential growth above a GI of 0.3-0.4. A similar model fitted to data adjusted for differences between regions and including soil P as a variable increased the proportion of variance accounted for from 59% t0 72%. The SW region was the only region for which differential growth of pioneer heather, taken across all years showed a significant non-linear response to GI compared to a linear model. For woody growth, pioneer heather showed a similar overall negative response to GI whilst the relationship was significantly quadratic for mature heather. Weight of woody material of building heather showed no overall response to variation in GI This study was only partially successful in describing relationships between Grazing Index and shoot productivity and weight of woody material. Contrary to expectations, pioneer heather was apparently less tolerant of grazing pressure than mature heather. The models suggested a stimulation of shoot growth on mature heather at low grazing pressure (<0.4 GI), and suppression at high, consistent with predictions of Palmer (1997). However the model was dependant on very few data points at higher GI levels, which also makes threshold identification uncertain. Pioneer heather generally showed a linear response of suppression with increasing grazing intensity. The apparent

		difference in susceptibility between pioneer and mature heather was supported by data for woody growth, variation in which may reflect differences in grazing pressure over a number of years. Across the whole data set heather shoot growth showed a positive correlation with weights of woody material. For mature heather it appeared that the largest and most productive plants tend to result from intermediate levels of grazing, but no such relationship was observed in young heather. In conclusion it was tentatively suggested that a grazing index of 0.4 may be optimum for mature heather, and tentatively suggested that this might represent a threshold for pioneer heather in some circumstances.
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	MAFF

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	d) Over what timescales can grazing-related change in plant structure and diversity be observed or expected? e) How is 'under-grazing' defined? What are the effects of low intensity regimes, set to restore small areas of priority habitat within a moorland mosaic, on other parts of the moorland including non-target habitats such as acid grassland? g) Do different types of livestock (species and breed), and combinations of livestock, affect moorland habitats differentially?

Study details	Authors	MITCHELL, R. J., ROSE, R. J. & PALMER, S. C. F.
	Year	2008
	Aim of study	To test the efficacy of different grazing regimes and intervention techniques aimed at establishing *Calluna vulgaris*
	Study design	2
	Quality score	++
	External validity	++
Population and setting	Source population	Sites typical of degraded upland moorlands within the UK

	Eligible population	2 contrasting grass-dominated moorland sites in England and Wales
	Inclusion and exclusion criteria	At both sites over-grazing resulted in decline in *Calluna* since 1970s
	Setting	*Nardus* site at Pwllpeiran; *Molinia* site at Redesdale
Methods of allocation to intervention/control	Methods of allocation	At *Nardus* (Pwllpeiran) site allocation was randomised and grazing regimes replicated. At *Molinia* (Redesdale) site grazing regimes were not replicated.
	Intervention description	September 2002: plots randomly assigned to one of 3 disturbance treatments: 'undisturbed'; 'rotavation'; 'trampling'. March 2003: plots had 2 sub-treatments applied: -*Calluna* seed on half of each plot; -No grazing (fencing) on half of each plot. Within each quarter plot, 1 4x4m sub plot established within which all recording carried out. Within each sub-plot, 9 1x1m permanent quadrats to record *Calluna* establishment and bare ground
	Control/comparison description	Sub-plots compared against one another. No control plot
	Sample sizes	*Nardus* (Pwllpeiran) site (dominated by *Nardus*, *Agrostis*, *Festuca*, with some *Vaccinium*): -3 blocks of land x 3 fields (5-7ha) in each block

		-each block, 3 fields randomly assigned to: 'cattle'; 'mixed'; 'sheep' -each field, 6 10x10m plots in areas with similar vegetation *Molinia* (Redesdale)site (dominated by *Molinia* with small amounts of *Calluna*): -3 fields (21-29ha) -'mixed high' (sheep and cattle) -'mixed low' (sheep and cattle) -'sheep only' -each field 18 10x10m plots Both sites 54 plots in total
	Baseline comparisons	NR
	Study sufficiently powered	NR
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Presence/absence of pioneer stage *Calluna* plants Cover of bare ground
	Secondary outcome measures	Seed-bank composition Effects of treatments on *Calluna* establishment: -heather seeding -disturbance treatments, bare ground -removing grazing -different grazing regimes
	Follow-up periods	Seed-bank sampled in Sep 2002 (before addition of seed) Grazing regimes started in spring 2003, continued until autumn 2006

	Methods of analysis	Separate analyses for each site. -Used generalized linear mixed models (GLMM) -'Fixed effect' both sites: disturbance, fencing, seeding, visit, and their interactions; *Nardus* site: block and grazing regime; *Molinia* site: field. -'Random-effect': plot -'Continuous variables': bare ground, *Calluna* morphology
Results		The addition of *Calluna* seed is a key factor in the establishment of *Calluna* on grass-dominated moorlands. The second key factor in the establishment of *Calluna* was the creation of small areas of bare ground for germination. The intensity (amount) of disturbance is important rather than the method used to create bare ground. When the disturbance intensity was low and little bare ground was created, grazing increased *Calluna* occurrence. Results suggest that disturbance doing more than just creating bare ground; may also limit re-growth of competitive grass species. At the *Nardus* site grazing by cattle only had equal or better *Calluna* establishment and growth than no grazing. Grazing by sheep alone was the least successful treatment for the *Nardus* sward. At the *Molinia* site, *Calluna* presence and the number of *Calluna* plants was greater in the grazed than the ungrazed plots; probably due to increased bare ground in grazed plots. However, *Calluna* plants in the grazed plots were much smaller; it is unlikely that these plants will grow into substantial *Calluna* bushes. This study showed that the creation of suitable sites for germination, the addition of

		Calluna seed and low intensity grazing are the key management tools for the establishment of *Calluna* on grass dominated swards.
Notes	Limitations identified by author	Cattle grazing (summer only) could be confounded with seasonality. Treatments carried out in small plots but were developed to be applicable at larger scales. The limiting factors are the economics of applying the initial treatment and thereafter maintaining appropriate stocking regimes. If implemented at a landscape scale, an increase in the use of cattle grazing – considerable change in farming practices in the British uplands.
	Limitations identified by review team	A longer period of time (5+years?) would have been preferable Limited number (2) of small sites (5-29ha) sites
	Evidence gaps and/or recommendations for further research	Further work needed to assess if these techniques work on the wide range of degraded grass-dominated heaths and moorland that occur throughout Europe. Further work to assess role of cattle in restoration projects; including cattle trampling as a one-off restoration treatment. Impact of nitrogen deposition on the success of restoration methods. Concerns over possible carbon release by any form of mechanical disturbance on peat-based soils. Conflicts between carbon conservation and habitat restoration will have to

		be addressed.
	Sources of funding	Defra, English Nature and CCW

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: _____Upland_____

Name of Review Sub-topic (if any): _____Moorland grazing_____

Review Question	d) Over what timescales can grazing-related change in plant structure and diversity be observed or expected? e) How is 'under-grazing' defined? What are the effects of low intensity regimes, set to restore small areas of priority habitat within a moorland mosaic, on other parts of the moorland including non-target habitats such as acid grassland? g) Do different types of livestock (species and breed), and combinations of livestock, affect moorland habitats differentially?
Study Citation	MITCHELL, R. J., ROSE, R. J. & PALMER, S. C. F. 2008. *Restoration of Calluna vulgaris on grass-dominated moorlands: the importance of disturbance, grazing and seeding*. Biological Conservation, 141, 2100-2111.
Study Design Category	2
Assessed by & when	Amy Christie 25-31/01/13

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described?	☐++	Comments: Sites typical of degraded upland moorlands within the UK. At both sites over-grazing resulted in decline in Calluna since 1970s
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e...g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Comments: 2 contrasting grass-dominated moorland sites in the UK: *Nardus* site at Pwllpeiran; *Molinia* site at Redesdale
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐+	Comments: Only 2 study sites, but these sites do typify many degraded moorlands throughout England and Wales Restoration of dwarf shrub was aim rather than specific NVC community. Intended that this broader aim would allow results to be applicable more widely within UK

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐+	Comments: At *Nardus* (Pwllpeiran) site allocation was randomised and grazing regimes replicated. This was not the case at *Molinia* (Redesdale) site. *Nardus* (Pwllpeiran) site (dominated by *Nardus, Agrostis, Festuca,* with some *Vaccinium*): -3 blocks of land x 3 fields (5-7ha) in each block -each block, 3 fields randomly assigned to: 'cattle'; 'mixed'; 'sheep' -each field, 6 10x10m plots in areas with similar vegetation *Molinia* (Redesdale)site (dominated by *Molinia* with small amounts of *Calluna*): -3 fields (21-29ha) -'mixed high' (sheep and cattle) -'mixed low' (sheep and cattle) -'sheep only' -each field 18 10x10m plots Both sites 54 plots in total
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: September 2002: plots randomly assigned to one of 3 disturbance treatments: 'undisturbed'; 'rotavation'; 'trampling'. March 2003: plots had 2 sub-treatments applied: -*Calluna* seed on half of each plot; -No grazing (fencing) on half of each plot. Within each quarter plot, 1 4x4m sub plot established within which all recording carried out. Within each sub-plot, 9 1x1m permanent quadrats to record *Calluna* establishment and bare ground. Details of materials and methods given – in sufficient detail to replicate.

2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐+	Comments: A longer period of time (5+years?) would have been preferable. Grazing regimes started in spring 2003, continued until autumn 2006.
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐++	Comments: Seed-bank sampled in Sep 2002 (before addition of seed). Only *Calluna* seedlings and <4year old plants recorded; mature and degenerate *Calluna* present prior to experiment discounted.
2.5 Were any other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (e.g. management not part of the experimental interventions, e.g. plots with unplanned burning)? Were groups treated equally?	☐++	Comments: Groups treated equally.
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐+	Comments: Only 2 sites but: Sites typical of degraded upland moorlands within the UK. At both sites over-grazing resulted in decline in Calluna since 1970s. Restoration of dwarf shrub was aim rather than specific NVC community. Intended that this broader aim would allow results to be applicable more widely within UK.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: Stocking rates fairly typical for this type of land (e.g. 0.5 cow/ha July and August; 1 – 1.5 ewes all year round). But cattle not always available on British upland farms. Disturbance treatments: 'undisturbed'; 'rotavation';

		'trampling' – probably reflect practice in many of these types of restoration projects.

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective? How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐+	Comments: Measurements: Presence/absence of pioneer stage *Calluna* plants - objective. Cover of bare ground assessed – relatively subjective No indication of QA
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++	Comments: Measurements appear completed as planned
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments: All important effects assessed: -Seed-bank composition -Effects of treatments on *Calluna* establishment: heather seeding disturbance treatments, bare ground removing grazing different grazing regimes
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐++	Comments: Yes, relevant: Direct measure of presence of pioneer *Calluna* within different treatments.
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐NA	Comments:
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Comments: Experiment ran for 4 years; a longer period of time (5+years?) would have been preferable.

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐++	Comments: Groups were very similar.
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐NR	Comments:
4.3 Were the estimates of effect size given or calculable?	☐NA	Comments:
4.4 Were the analytical methods appropriate? Were any important differences in post-treatment time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐++	Comments: Separate analyses for each site. -Used generalized linear mixed models (GLMM) -'Fixed effect' both sites: disturbance, fencing, seeding, visit, and their interactions; *Nardus* site: block and grazing regime; *Molinia* site: field. -'Random-effect': plot -'Continuous variables': bare ground, *Calluna* morphology
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	Comments: p-values quoted throughout results section.
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study	☐++	Comments: Study design appears unbiased. Potential confounder – residual seed-bank was thoroughly investigated. Acknowledged that cattle grazing (summer only) could

design?		be confounded with seasonality.
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐++	Comments: Treatments carried out in small plots but were developed to be applicable at larger scales. The limiting factors are the economics of applying the initial treatment and thereafter maintaining appropriate stocking regimes. If implemented at a landscape scale, an increase in the use of cattle grazing – considerable change in farming practices in the British uplands.

Evidence Table

Name of Evidence Review:	Upland
Name of Review Sub-topic (if any):	Grazing
Review Question	f. what factors influence spatial patterns of grazing

Study details	Authors	Oom et al.
	Year	2010
	Aim of study	To identify whether increased heterogeneity of grass:dwarf-shrub mosaics leads to increased resistance to herbivory
	Study design	Quantitative experimental
	Quality score	+
	External validity	+
Population and setting	Source population	Upland moorland
	Eligible population	Calluna –vaccinium & f. ovina-a. capillaris-g.saxatile mosaic
	Inclusion and exclusion criteria	N/A
	Setting	NE Scotland
Methods of allocation to intervention/control	Methods of allocation	Six plots, three differing stocking rates (replicated twice). Method of allocation not reported

	Intervention description	Three stocking rate treatments (2, 3 & 4 sheep/ha) – all stocked with 6 sheep for varying lengths of time.
	Control/comparison description	Comparison between three stocking rates
	Sample sizes	Total of 675 transects and 3504 samples across the 6 plots
	Baseline comparisons	N/A
	Study sufficiently powered	+
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Calluna defoliation (% current shoot removal estimated) recorded at boundary between calluna and grass (track or patch)
	Secondary outcome measures	N/A
	Follow-up periods	Measurements taken twice per year Oct 1998-April 2001
	Methods of analysis	REML/Linear regressions
Results		Significant overall increase in heather defoliation with increasing stocking rate ($p=0.003$) Frequency and severity of defoliation higher for whole year than for summer only – showing defoliation continued though year Defoliation higher around grass patches than paths ($p<0.001$) and higher closer to edge of paths and grass patches than further away ($p<0.001$). Defoliation higher uphill of grass patch than downhill ($p<0.001$), although this association was found only in summer and not across a year as a whole.

		Defoliation is not always negatively correlated with distance from the grass edge, but is dependent on the spatial configuration of the mosaic (p=0.032) The contrast between edge and distant defoliation decreases at higher grazing pressures as impact zones are increasingly likely to overlap.
Notes	Limitations identified by author	Findings need to be verified at other sites
	Limitations identified by review team	Flock size kept constant in experiment but number of stocking days varied – different results may be recorded by varying pressure by changing flock size Growth stage of calluna as a variable was not considered
	Evidence gaps and/pr recommendations for further research	Replication on other sites/larger scale experiments
	Sources of funding	Not reported

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: _____UPLAND_____

Name of Review Sub-topic (if any): _____GRAZING_____

Review Question	f. what factors influence spatial patterns of grazing
Study Citation	Oom et al. 2010
Study Design Category	Quantitative experimental 2
Assessed by & when	SUSANNA PHILLIPS 04/10/2012

Section 1: Population		
1.1 Are **the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Location (NE Scotland)/altitude/aspect described NVC communities – spp composition described but not habitat condition
1.2 Are **the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: How plots located within study site not described Vegetation homogeneity between blocks not described Assumed plots representative of study site, but not clear
1.3 Are **the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Sampling lines located with randomised spacing Sampling at every grass-dwarf shrub interface Transects located subjectively perpendicular to grass-dwarf shrub boundary

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐++ ☐+ ☐- ☒NR ☐NA	Comments: Allocation of stocking rate to plot not described Stocking rate replicated x2
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☒++ ☐+ ☐- ☐NR ☐NA	Comments: Treatments sufficiently detailed to allow replication. Comparisons appear appropriate
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐++ ☒+ ☐- ☐NR ☐NA	Comments: No deviation from methodology recorded – assume stocking rate per plot as described
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐++ ☒+ ☐- ☐NR ☐NA	Comments: Not clear from way study reported
2.5 Were any other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions eg plots with unplanned burning)? Were groups treated equally?	☐++ ☐+ ☐- ☒NR ☐NA	Comments: No other intervention recorded

2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Representative of calluna- vaccinium heath & agrostis-festuca-galium saxatile grasslands
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Assumes total livestock removal dec-feb is standard agricultural practice, results may not be valid for winter grazed moors

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Heather defoliation – shoots assigned to category by % grazing = subjective assessment, but no QA/validation recorded
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Data suggests all measurements completed
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Outcomes meet objectives of study

3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Direct measurement of grazing on calluna
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Defoliation recorded for summer in oct/whole year in apr – same approach across all 3 interventions
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: 3 year experiment

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Assumed all groups have no grazing on current seasons shoots at start of each year
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: 675 transects 3504 samples

4.3 Were the estimates of effect size given or calculable?	☐++ ☐+ ☐- ☐NR ☐NA	Comments:
4.4 Were the analytical methods appropriate? Were any important differences in post-treatment time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Unbalanced design – REML Linear regressions
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: P-values given
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐++ ☐+ ☐-	Comments: Not reported if validated for observer error
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐++ ☐+ ☐-	Comments: Assumption that livestock removal dec-feb is standard agricultural practice – not universally applicable - may significantly alter findings. Appropriate for sites with winter livestock removal.

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?

Study details	Authors	Oom SP, Sibbald AM, Hester AJ, Miller DR, Legg CJ
	Year	2008 (study done 1998-2001)
	Aim of study	Impacts of sheep grazing a complex vegetation mosaic: Relating behaviour to vegetation change.
	Study design	RCT
	Quality score	=QA5.1 Heather defoliation measurements were averaged over all seasons and years, providing a single estimate of edge-heather defoliation for each of the grass patches sampled Sources of error for the aerial photography described in Oom, 2003 (unpublished PhD thesis, University of Edinburgh)
	External validity	=QA5.2 North east Scotland and Scottish Blackface sheep not necessarily transferable to English uplands
Population and setting	Source population	Glenshaugh Research Station, northeast Scotland. Altitude200-250m. NNW facing Calluna dominated moorland
	Eligible population	Highly fragmented mosaic, predominately Calluna with numerous patches of Agrostis/Festuca grass.
	Inclusion and exclusion criteria	

	Setting	
Methods of allocation to intervention/control	Methods of allocation	6x1ha plots. Three grazing treatments
	Intervention description	3 grazing treatments of 4,3,2& 1-year-old Scottish Blackface sheep per 0.1ha randomly allocated over 6x1ha plots Site left ungrazed for 1 year prior to this study but grazed for 8 weeks/year from 1991 with sheep and red deer. 3 grazing treatments of 4,3,2& 1-year-old Scottish Blackface sheep per 0.1ha randomly allocated to plots 1&5, plots 2&6 and plots 3&4 between March and November each year. Groups of 6 animals were put in each of the plots for different numbers of days in a 3-week rotational schedule to keep the required stocking rates while keeping animal densities the same. Heather defoliation measured in October and April each year at the beginning and end of the grazing season Vegetation change was measured from aerial photographs taken at beginning (October 1998) and end(October 2001) of the experiment Heather defoliation recorded using 7x100m sampling lines in each plot, positioning minutely described Each sheep individually fleece-marked for distance id. Behaviour categories used were foraging and resting (lying) only. Vegetation change was measured by aerial photography at beginning and end and image classification
	Control/comparison description	
	Sample sizes	
	Baseline comparisons	
	Study sufficiently powered	
Outcomes and methods of analysis (inc effect	Primary outcome measures	Changes in vegetation cover related to sheep behaviour Changes in spatial pattern of vegetation related to sheep behaviour

size, CIs for each outcome and significance)		
	Secondary outcome measures	Effects of heather defoliation Effects of sheep resting behaviour
	Follow-up periods	Carried out over a period of 3 years
	Methods of analysis	
Results		Changes in vegetation cover – net change over all plots from shrub to mixed vegetation and to a lesser extent to degraded heather. Changes in spatial pattern of vegetation – small changes were spread fairly evenly across the mosaic. Larger changes, particularly those from shrub to mixed vegetation, were concentrated in a few areas Effects of heather defoliation- there were significant linear relationships between heather defoliation rate and the % decrease in shrub vegetation/increase in mixed vegetation associated with adjacent grass patches. However, relationship[s between defoliation rate and percentage changes in grass and degraded heather were not significant. Effects of sheep resting behaviour – the combined increase in grass and mixed vegetation cover at resting sites was equivalent to an increase of 27% by area of the grass patches originally classified as resting sites at the start of the experiment. 'Spatially aggregated patterns of behaviour (i.e. resting) clearly played an important role and would have been driven, in part, by the initial spatial patterns of vegetation in the different plots.'
Notes	Limitations identified by author	No replicate images of the remote sensing detailed vegetation maps were available except at the beginning and the end of the experiment, preventing a rigorous error analysis

	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	Macaulay Development Trust, Scottish Executive Environment and rural Affairs Department

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: _____Upland _____

Name of Review Sub-topic (if any): _____Moorland grazing_____

Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?
Study Citation	Oom SP, Sibbald AM, Hester AJ, Miller DR, Legg CJ .Impacts of sheep grazing a complex vegetation mosaic: Relating behaviour to vegetation change. Agriculture, Ecosystems and Environment 124 (2008) pp 219-228
Study Design Category	1
Assessed by & when	Alison Hiles 8/2/2013

Section 1: Population		
1.1**Are the source population(s) or area(s) well described?** e. g. Were habitat(s) and biodiversity of the area(s) well described.	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: Glenshaugh Research Station, northeast Scotland. Altitude200-250m. NNW facing Calluna dominated moorland
1.2**Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e. g. is the floristic diversity representative of the habitat? Were important groups under-represented?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: Highly fragmented mosaic, predominately Calluna with numerous patches of Agrostis/Festuca grass.
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: 6x1ha plots. Three grazing treatments

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: 3 grazing treatments of 4,3,2& 1-year-old Scottish Blackface sheep per 0.1ha randomly allocated over 6x1ha plots Site left ungrazed for 1 year prior to this study but grazed for 8 weeks/year from 1991 with sheep and red deer.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: 3 grazing treatments of 4,3,2& 1-year-old Scottish Blackface sheep per 0.1ha randomly allocated to plots 1&5, plots 2&6 and plots 3&4 between March and November each year. Groups of 6 animals were put in each of the plots for different numbers of days in a 3-week rotational schedule to keep the required stocking rates while keeping animal densities the same. Heather defoliation measured in October and April each year at the beginning and end of the grazing season Vegetation change was measured from aerial photographs taken at beginning (October 1998) and end(October 2001) of the experiment Heather defoliation recorded using 7x100m sampling lines in each plot, positioning minutely described Each sheep individually fleece-marked for distance id. Behaviour categories used were foraging and resting (lying) only. Vegetation change was measured by aerial photography at beginning and end and image classification
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: 3 year study
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or	☐++ ☐+ ☐-	Comments:

vice versa? Was it sufficient to cause important bias?	☐NR ✔ ☐NA	
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions eg plots with unplanned burning)? Were groups treated equally?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++ ✔ ☐+ ☐- ☐NR ☐NA	Comments: North east Scotland and Scottish Blackface sheep not necessarily relevant to English uplands
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++ ✔ ☐+ ☐- ☐NR ☐NA	Comments: sheep are not usually fenced on heather moorlands

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++ ✔ ☐+ ☐- ☐NR ☐NA	Comments: Outcomes were measured and a linear regression analysis was applied.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined	✔ ☐++ ☐+ ☐-	Comments:

study outcome definitions)?	☐NR ☐NA	
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments:
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐++ ☐+ ☐- ☐NR	Comments:

	✔ ☐NA	
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐++ ☐+ ☐- ✔ ☐NR ☐NA	Comments:
4.3 Were the estimates of effect size given or calculable?	☐++ ☐+ ☐- ✔ ☐NR ☐NA	Comments:
4.4 Were the analytical methods appropriate? Were any important differences in post-treatment time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study	☐++ ✔ ☐+ ☐-	Comments: Heather defoliation measurements were averaged over all seasons and years, providing a single estimate of edge-heather defoliation for each of the grass patches sampled Sources of error for the aerial photography described in Oom, 2003 (unpublished PhD thesis, University of Edinburgh)

design?		
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐++ ✔ ☐+ ☐-	Comments: North east Scotland and Scottish Blackface sheep not necessarily transferable to English uplands

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	What are the effects of grazing regimes and stocking rates on the maintenance and or restoration of moorland biodiversity and ecosystem service delivery? a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?

Study details	Authors	Pakeman, R.J., Hulme, P.D., Torvell, L. & Fisher, J.M.
	Year	2003
	Aim of study	To investigate the suitability of different grazing treatments for rehabilitating degraded dry heath, and derive grazing management prescriptions
	Study design	Replicated block with four treatments applied randomly? And additional sheep and sheep+rabbit exclusion blocks.
	Quality score	+
	External validity	+
Population and setting	Source population	Upland dwarf shrub heath
	Eligible population	Area of degraded heather moorland H12 *Calluna-Vaccinium*
	Inclusion and exclusion criteria	Site of blocks chosen in area of fairly uniform grass/ heath vegetation.

	Setting	Moorland site in Morayshire, NE Scotland. 300m AOD
Methods of allocation to intervention/control	Methods of allocation	Treatment blocks selected (subjectively?) after a preliminary survey of the site. Proportions of grass and dwarf shrub stated to be similar across the chosen area. Each block has four treatment plots, with smaller nested area fenced from both sheep and rabbits. Block has a further two areas outside the treatments fenced against sheep.
	Intervention description	Four grazing treatments – winter and summer high and low, in addition to areas fenced against sheep, and sheep and rabbits.
	Control/comparison description	Comparison is the typical background grazing levels of the hillside outside of the blocks.
	Sample sizes	Two replicates per treatment, with one sheep/rabbit exclosure in each treatment block. Heather utilisation measured from 100 sample points per block, sward heights and *Nardus* utilisation from 40 points per block. Species frequency measured from 5 pins at 20 points per grazing treatment block. Same total number of points collected over the sheep and sheep/ rabbit exclosures in each block.
	Baseline comparisons	Measurements made in fist year when plots were fenced. In the unfenced area recording began in 1992
	Study sufficiently powered	No power analysis reported. In these type of experiments sample size (i.e. number of treatments and replication) needs to be balanced against cost and practicality.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Utilisation rates of heather and mat grass. Frequency of each species and structural components of heather. Sward heights.
	Secondary outcome measures	none

	Follow-up periods	Treatments in place for 5 years, measurements made each year except unfenced area.
	Methods of analysis	Repeated measures analysis of variance for utilisation and height, and ordination techniques for floristic data. Principal Response Curves used to demonstrate separation of each treatment effect from the year round high grazing control.
Results		All fenced treatments showed an increase in heather frequency over 5 years of the experiment, with the increase in proportion to the reduction in stocking rate, irrespective of timing. There is a resultant decline in utilisation as heather increases in frequency. Other dwarf shrub species also benefitted from reduced grazing. Declines were observed in the grass *Agrostis capillaris* and total monocotyledonous (grass and related) Species. Only small differences were observed between winter or summer low rates and no sheep grazing treatments. Comparison of dwarf shrub recovery in the sheep and sheep/rabbit exclosures show measurable rabbit effect on recovery of heather. Principal Response Curves demonstrate the separation of each treatment from the year-round heavily grazed control area. On this dry heath system a reduction in sheep numbers to 0.8/0.9 sheep ha^{-1} yr^{-1} to give utilisation levels below 20% was seen to achieve the desired result of improving vegetation condition. There was little effect of timing of grazing. It is suggested however that measured utilisation of heather provides a better basis for setting grazing management than sheep numbers.
Notes	Limitations identified by author	Controlled grazing experiments do not fully simulate the effect of the same management on open moorland. In particular there is an opportunity for heather re-growth between periods of grazing.
	Limitations identified by review team	The authors suggest above limitation is largely overcome through comparing utilisation rates rather than stocking rates. However guide average stocking rates are presented but the basis for deriving these, from short bursts of grazing on small plots, may not readily scale up.
	Evidence gaps and/pr	Identification of the threshold cover of dwarf shrub below which it is necessary to

	recommendations for further research	actively restore dwarf shrub, and above which grazing manipulation can achieve results (DM note: This may also depend on the target heather cover we want to achieve and degree of mosaic). Also how this figure is influenced by the age of heather, other species present, and environmental conditions.
	Sources of funding	SEERAD

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: _____Uplands_____

Name of Review Sub-topic (if any): ____Moorland Grazing_____

Review Question	
Study Citation	Pakeman, R.J., Hulme, P.D., Torvell, L. & Fisher, J.M. (2003) Rehabilitation of degraded dry heather *Calluna vulgaris* moorland by controlled sheep grazing.
Study Design Category	
Assessed by & when	D Martin 15/10/12

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐+	Comments: The heather moorland population is not described in great detail, but distribution and significance, and context – loses- are described.
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++	Comments: location and closest NVC community given. The site is typical of one of the more extensive upland heathland communities in the UK.
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐+	Comments: Treatment blocks selected (subjectively?) after a preliminary survey of the site. Proportions of grass and dwarf shrub stated to be similar across the chosen area.

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐+	Comments: It is not reported that treatments were randomised, but analysis methods suggest this was the case. If vegetation and soil type is relatively uniform there should be minimal confounding effects. Rabbit grazing is also controlled for.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐-	Comments: Treatments are well described, including actual numbers and timing of livestock, so easily replicable. However, the treatments are not easy to relate to actual practice. The impacts of short periods of grazing may not be representative of the stated annual sheep equivalent, if that rate were derived from a year-round grazing regime. It is not clear initially what the target utilisation rates were.
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐+	Comments: Five years of grazing treatment – this is a reasonable length exposure for such a study and provides enough time for responses to develop, although further change would be likely over a longer timescale. First year of grazing on summer treatments was higher than subsequent years as the target utilisation rates were sought.
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐+	Comments: No evidence that plots were grazed at the wrong time, or that ungrazed plots received grazing. However there is a suggestion that the sheep exclusion plots may have allowed increased rabbit grazing, although there is still a difference between the sheep and sheep/ rabbit exclusion plots.
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐++	Comments: No evidence of other interventions
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Comments: Site appears fairly representative of typical upland dry heather moorland found throughout upland areas of Scotland, N England and Wales.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK	☐-	Comments: Some concerns that the short periods of grazing, and winter only treatments, do not reflect

practice(s)?		typical practice, and perhaps not so readily translatable to typical practice as suggested.

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐+	Comments: Measured variables were utilisation rates of heather and mat grass, sward heights and species frequency. They all follow standard methods designed to minimise observer error, e.g. use of HFRO sward stick to measure height, and point quadrats, placed at permanently marked points on initially random transects. There is an element of subjectivity involved in estimating heather utilisation. First estimate carried out in May – this may be unreliable due to new season's growth.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐+	Comments: All measurements completed annually in the field for treatments. The unfenced control was only surveyed from 1992 onwards (1990 for treatments)
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐+	Comments: Main outcomes relating to the aims of the study were assessed. Perhaps utilisation of other key species (e.g. *Vaccinium myrtillus*) could have been estimated.
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐++	Comments: Variables all directly relevant to the aims of the study.
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Comments: Treatments all applied for same number of years, and measurements made on all treatments annually
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Comments: Five years of grazing treatment – this is a reasonable length exposure for such a study and provides enough time for responses to develop – in part of the analysis treatment effects only became significant in 1994. In a dynamic system further change would be likely over a longer timescale.

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]?	☐+	Comments: Yes, the study area was said to be roughly uniform in terms of proportion of grass to dwarf shrub and low occurrence of herbs across the site.

Were there any differences between groups in important confounders at baseline?		
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐NR	Comments: No power analysis reported. In these type of experiments sample size (i.e. number of treatments and replication) needs to be balanced against cost and practicality.
4.3 Were the estimates of effect size given or calculable?	☐-	Comments: No estimate given, but a range of treatments tested.
4.4 Were the analytical methods appropriate? Were any important differences in post-treatment time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐++	Comments: Randomised block ANOVA on derived coefficients from utilisation and sward height figures, with control treated separately due to fewer years observations. Ordination of floristic data, again adjusted for the lack of early data in the control (average of other treatments for 1990, all 1991 treatment data dropped)
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++	Comments: p values generally given for analysis of variance results for utilisation and heights, and time and treatment effects on species frequency. Significance of treatment response within the ordination is also given
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐+	Comments: Experiment well designed and controlled, and treatments applied consistently. Site reasonably uniform, but likely to be some variation. Only two replicates.
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐+	Comments: Habitat is widespread and site likely to be representative. Some concern over how treatments relate to actual year-round grazing regimes on moorland.

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	b) What methods of stocking rate calculation, or setting grazing regimes, consistently provide regimes that maintain or restore moorland biodiversity, and what are the key parameters that calculations should include?

Study details	Authors	Pakeman, R.J & Nolan, A.J.
	Year	2009
	Aim of study	To identify a utilization level below which heather shout increase in cover at the expense of monocots, and to calculate the associated levels of uncertainty. To compare the findings to an analysis based on stocking rate data.
	Study design	2 Multi site analysis of block treatment experiments (each site with one or two replicates).
	Quality score	++
	External validity	++
Population and setting	Source population	Upland heathland heather moorland. Some of the conservation interest briefly described.
	Eligible population	The ten sample areas cover three upland heather moorland areas in Scotland and north-east England, so likely to be representative of a range of upland heathland habitats and mosaics. They are reported as covering a wide range of situations in terms

		of starting heather cover, community composition and growth phase.
	Inclusion and exclusion criteria	Heathland habitat, sheep grazing treatments, utilisation and cover measured consistently by accepted methods.
	Setting	Sites in West Scotland; Argyle and West Highlands; North-east Scotland and Northumberland NE England.
Methods of allocation to intervention/control	Methods of allocation	Method of allocation is not covered on the brief descriptions of each experiment. Not enough detail presented here on each site to identify whether there were potential sources of bias (randomised?).
	Intervention description	The treatments are controlled stocking rates from 0.4 to 2.1 sheep per ha per year, at three levels per site and/ or combinations of summer or year round grazing. Also stock exclusion or open hill grazing treatments. At one site (Glensaugh), treatments applied in terms of utilisation rates with the highest (year round 80%) equating to 6 sheep er ha.
	Control/comparison description	This meta-analysis as a correlative study. The individual experiments usually have moderate or high treatments that are comparable with typical open hill stocking rates, and low or moderate treatments that are similar to conservation grazing regimes. Where open hill grazing rate is recorded, it is similar to high treatments in some experiments (1.8 ewes per ha per year).
	Sample sizes	Ten experiments of up to 6 treatments, with up to two replicates. Seventy-four plots in total with 100 utilisation measurements each and vegetation measurements in each plot.
	Baseline comparisons	Not reported in detail, but stated that starting points were different, although may have been similar within an experiment. However the experiments were chosen to cover a range of starting points.

	Study sufficiently powered	Not reported, but multiple sites likely to increase power to detect a real relationship. Seventy-four plots in total.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Outcome measurements are objective measures of utilisation, using same technique in all experiments, by estimating the proportion of current year's growth removed. This is a reasonably robust technique, but some subjectivity in estimating the proportion of shoot removed. Vegetation change also measured over the course of the experiments, using point quadrats.
	Secondary outcome measures	
	Follow-up periods	Most experiments in place for 5-6 years
	Methods of analysis	Linear regression of rate of change in proportion of heather against time. Fitted against mean utilisation or sheep stocking density using linear mixed-effects model with residual maximum likelihood as the fit criterion. Utilisation/ stocking density as fixed effect and other variables added in turn.
Results		From analysis of the ten experiments there was a clear relationship between rate of change in the proportion of heather and its utilisation. Additional terms tried in the model including season, region and growth phase were not significant. A utilization level of 31.6% of current season's growth was found to maintain the balance between heather and monocots. However the 95% confidence intervals for no change are 22.5% and 41.4%, indicating a considerable degree of variation and uncertainty. There is a similar relationship between stocking rates and utilisation, although this variable explains slightly less of the variance. No change is expected at 1.82 sheep ha^{-1} yr^{-1}, with 95% confidence intervals at 1.14 and 2.61 sheep ha^{-1} yr^{-1}. As most of the sites were established on a mix of heather and grass, the no effect stocking rate may be higher than can be withstood where heather has higher cover with less grass.

		The relationship between stocking rate and percentage utilisation was strong, with an increase of 15.1% for each 1 sheep ha^{-1} yr^{-1}. The current assumed sustainable utilization level of 40% may be too high and it is recommended it is set nearer 20% to reduce the risk of heather loss (where this is an objective). It is suggested that developing models based on utilisation data is more efficient than using stocking rate, which needs to take more account of different vegetation types.
Notes	Limitations identified by author	The experiment does not encompass sites with higher cover of heather which may sustain lower stocking rates. Possible limitations of using rate of change in proportion of heather as a measure, as it does not take account of overall productivity and grazing preferences. However it can reflect a range of responses to different grazing scenarios.
	Limitations identified by review team	Few limitations of methodology or analysis. Utilisation rates is a useful tool, but still requires monitoring of utilisation in different parts of the moor and adjustment of stocking rates. Doesn't tell us anything about impacts on other components of heathland or co-located habitats.
	Evidence gaps and/pr recommendations for further research	Additional sites may reduce the spread of confidence intervals.
	Sources of funding	Scottish Government Rural and Environmental research and Analysis Directorate. Original experiments funded by different sources.

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: _____Upland _____

Name of Review Sub-topic (if any): _____Moorland Grazing_____

Review Question	b) What methods of stocking rate calculation, or setting grazing regimes, consistently provide regimes that maintain or restore moorland biodiversity, and what are the key parameters that calculations should include?
Study Citation	Pakeman, R.J. & Nolan, A.J. (2009) Setting sustainable grazing levels for heather moorland: a multi-site analysis. Journal of Applied Ecology 46, 363-368
Study Design Category	1
Assessed by & when	D Martin 6/11/12

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐+	Comments: Upland heathland heather moorland. Some of the conservation interest briefly described.
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++	Comments: The ten sample areas cover three upland heather moorland areas in Scotland and north-east England, so likely to be representative. Many of experiments are unpublished but three are published and reviewed separately as well as for this multi-site paper. (Pakeman *et al*, 2003; Grant *et al*, 1978; Hulme *et al*, 2002). They are reported as covering a wide range of situations in terms of starting heather cover, community composition and growth phase.
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐+	Comments: The experimental plots are on vegetation described as wet heath (M15 heather – cross-leaved heath- deer grass) or dry heath (H12 heather-bilberry heath). These are widespread upland communities and the experimental plots are likely to have been chosen to reflect the vegetation of the wider site, although likely to have been subjective. Site and experiment details are appended

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐NR	Comments: Method of allocation is not covered on the brief descriptions of each experiment. Not enough detail presented here on each site to identify whether there were potential sources of bias (randomised?).
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐+	Comments: The treatments are quite well described in terms of stocking rates or utilisation levels (though not always- Otterburn), but open hill grazing rates, where this was a treatment, not always reported. Seasonal treatments identified but start and end dates not given.
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐++	Comments: Where reported the length of exposure to treatments is broadly similar at 5-6 years. Not reported at two sites. These durations would be adequate to obtain reliable utilisation results. Most experiments took place during early-mid 1990s, but one in 1970s.
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐NR	Comments:
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐NR	Comments:
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Comments: Yes – a range of upland sites through Scotland and n England.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: Grazing treatments often moderate or high treatments that are comparable with typical open hill stocking rates, and low or moderate treatments that are similar to conservation grazing regimes. Where open hill grazing rate is recorded, it is

		similar to high treatments in some experiments (1.8 sheep ha^{-1} yr^{-1}). Small plot treatments do not necessarily reflect behaviour on the open hill, however.

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐+	Comments: Outcome measurements are objective measures of utilisation, using same technique in all experiments, by estimating the proportion of current year's growth removed. This is a reasonably robust technique, but some subjectivity in estimating the proportion of shoot removed. Vegetation change also measured over the course of the experiments, using point quadrats.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐+	Comments: Assumed that all measurements complete. Experiments included as they are consistent and provide enough suitable data for inclusion.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments: For the purposes on the multi-site meta analysis, only utilisation looked at as this has been done consistently across experiments.
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐++	Comments: Yes – well defined and accepted techniques for estimating utilisation.
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐+	Comments: Most experiments ran for 5-6 years, similar duration. Most experiments took place during early-mid 1990s, but one in 1970s
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++	Comments: Yes – generally 5-6 years

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they	☐-	Comments: Not reported in detail, but stated that starting points were different, although may have

adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐	been similar within an experiment. However the experiments were chosen to cover a range of starting points.
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	Comments: Not reported, but multiple sites likely to increase power to detect a real relationship.
4.3 Were the estimates of effect size given or calculable?	☐-	Comments: Not presented. The multi-site analysis is a correlative study.
4.4 Were the analytical methods appropriate? Were any important differences in post-treatment time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐++	Comments: Linear regression of rate of change in proportion of heather against time. Fitted against mean utilisation or sheep stocking density using linear mixed-effects model with residual maximum likelihood as the fit criterion. Utilisation/ stocking density as fixed effect and other variables added in turn.
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++	Comments: 95% confidence intervals calculated for tests of effect of variables in the model. P values presented for regressions.
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐++	Comments: multi-site study, with replicates at some sites. This will increase the power to detect a relationship over individual studies.
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐++	Comments: A range of sites included, across different geographical areas and starting points of vegetation composition including heather cover.

Grant, S.A., Barthram, G.T., Lamb, W.I.C. & Milne, J.A. (1978). Effects of season and level of grazing on the utilization of heather by sheep. 1. Responses of the sward. *Journal of the British Grassland Society*, 33, 311-320

Hulme, P. D., MERRELL, B. G., TORVELL, L., FISHER, J. M., SMALL, J. L. & PAKEMAN, R. J. (2002). Rehabilitation of degraded Calluna vulgaris (L.) Hull-dominated wet heath by controlled sheep grazing. *Biological Conservation,* 107**,** 351-363.

Pakeman, R. J., Hulme, P. D., Torvell, L. & Fisher, J. M. 2003. Rehabilitation of degraded dry heather Calluna vulgaris (L.) Hull moorland by controlled sheep grazing. *Biological Conservation,* 114**,** 389-400.

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	Moorland Grazing and stocking rates

Study details	Authors	Stephen C. F. Palmer, Alison J. Hester, David A. Elston, Iain J. Gordon, and Sue E. Hartley, Journal of Ecology
	Year	2003
	Aim of study	To examine how the distributions of grass (preferred vegetation) within a heather-dominated (less-preferred vegetation) landscape influence spatial variation in heather utilisation by free-ranging red deer and sheep at a range of spatial scales.
	Study design	Designed to examine the impact of distribution of grass on the utilisation of heather by herbivores. 0.25km squares were used, containing either 1-8% grass or >12% grass Type 2
	Quality score	=QA 5.1 No adjustment made for potential supplementary feeding, location of water points, disturbance, changes in weather etc. and their effects on ranging behaviour. +
	External validity	=QA 5.2 Would need to be more sheep-orientated to be of relevance in England+
Population and setting	Source population	6 Land management units in the Cairngorms
	Eligible population	Selected by red deer density – 3xlow, 2x medium and 3xhigh No account taken of sheep, mountain hare or rabbit density
	Setting	Designed to examine the impact of distribution of grass on the utilisation of heather by herbivores. 0.25km squares were used, containing either 1-8% grass or >12% grass

Methods of allocation to intervention/control	Methods of allocation	Selected by red deer density – 3xlow, 2x medium and 3xhigh
	Intervention description	NA
	Control/comparison description	Designed to examine the impact of distribution of grass on the utilisation of heather by herbivores.
	Sample sizes	0.25km squares were used, containing either 1-8% grass or >12% grass
	Baseline comparisons	NA
	Study sufficiently powered	NR
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Measurements accurately taken of heather heights in 'grass-patch edge' and 'distant zones'
	Secondary outcome measures	Greatest proportions of grass in the area led to greater heather grazing.
	Follow-up periods	Autumn 1998 and spring 1999
	Methods of analysis	Within each square, heather utilisation was measured along 8 transects, which were distributed around the square to make up a representative sample of the grass patch sizes and dominant grass species. 12 sample quadrats were placed along each transect
Results		Grass attracts grazers. The main effect on utilisation at quadrat level was distance from the grass patch edge. There was a sharp decline in utilisation with increased distance from grass in all the land management units. On all units, heather was much shorter within 1-2m of the grass patch edge than further away. Grass availability did not significantly affect relative heather heights in the edge and distant zones of the transects

		However, the dominant grass species did affect heather height. Agrostis/Festuca patches showed the greatest proportional height difference at 32%, followed by Nardus at 26% and then Molinia at 11%. This demonstrates that the heather receives much higher impact when adjacent to preferred grass vegetation.
Notes	Limitations identified by author	
	Limitations identified by review team	No adjustment made for potential supplementary feeding, location of water points, disturbance, changes in weather etc. and their effects on ranging behaviour.
	Evidence gaps and/pr recommendations for further research	Work on the effects of sheep rather than deer
	Sources of funding	Macaulay Institute

Quality Assessment Checklist: Quantitative Study Observational / Correlation v2.0

Name of Evidence Review: _____Upland_____

Name of Review Sub-topic (if any): ____Moorland grazing_____

Review Question	Moorland Gazing and stocking rates
Study Citation	The Perils of having tasty neighbours:grazing impacts of large herbivores at vegetation boundaries. Stephen C. F. Palmer, Alison J. Hester, David A. Elston, Iain J. Gordon, and Sue E. Hartley, Journal of Ecology
Study Design Category	1
Assessed by & when	Alison Hiles 28/1/13

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐++ ✔ ☐+ ☐- ☐NR ☐NA	Comments: 6 land management units in the Cairngorms. 3 in eastern half and 3 in western half.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++ ☐+ ✔ ☐- ☐NR ☐NA	Comments: Selected by red deer density – 3xlow, 2x medium and 3xhigh No account taken of sheep, mountain hare or rabbit density
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++ ✔ ☐+ ☐- ☐NR ☐NA	Comments: Designed to examine the impact of distribution of grass on the utilisation of heather by herbivores. 0.25km squares were used, containing either 1-8% grass or >12% grass

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐++ ✔ ☐+ ☐- ☐NR ☐NA	Comments: Designed to examine the impact of distribution of grass on the utilisation of heather by herbivores. 0.25km squares were used, containing either 1-8% grass or >12% grass
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++ ✔ ☐+ ☐- ☐NR ☐NA	Comments:
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐++ ☐+ ☐- ✔ ☐NR ☐NA	Comments: Difficult to tell. Free-ranging herbivores
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:
2.5 Is the setting applicable to the UK?	☐++ ☐+ ✔ ☐- ☐NR ☐NA	Comments: Not really to England. Grazing by populations of deer with few sheep is not really comparable with the sheep-dominated fells in England

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐++ ✔ ☐+ ☐- ☐NR ☐NA	Comments: Measurements accurately taken of
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐++ ✔ ☐+ ☐- ☐NR ☐NA	heather heights in 'grass-patch edge' and 'distant zones' (this last line and a bit belongs in the box above but I can't move it up)
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++ ✔ ☐+ ☐- ☐NR ☐NA	Comments: 'The strong effect of distance was expected from paddock studies' Also greatest proportions of grass in the area led to greater heather grazing. Grass attracts grazers.
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:
3.5 Were there similar follow up times in exposure and comparison groups?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:

3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐++ ☐+ ✔ ☐- ☐NR ☐NA	Comments: The study only covered one autumn and one spring measurement 1998-1999

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐++ ☐+ ☐- ✔ ☐NR ☐NA	Comments:
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐++ ☐+ ☐- ✔ ☐NR ☐NA	Comments:
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐++ ☐+ ✔ ☐- ☐NR ☐NA	Comments: Study not long-term enough to adjust for confounders. Stats techniques included Anova, ordination and regression analysis
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments: Herbivores free-ranging over only one year

Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐++ ☐+ ✔ ☐-	Comments: No adjustment made for potential supplementary feeding, location of water points, disturbance, changes in weather etc. and their effects on ranging behaviour. DM+ Well designed, and analysis accounts for reduced independence from restricted random sampling
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐++ ☐+ ✔ ☐-	Comments: Would need to be more sheep-orientated to be of relevance in England DM+ Well designed, and analysis accounts for reduced independence from restricted random sampling

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?

Study details	Authors	Pearce-Higgins, J. W. & grant, M. C
	Year	2006
	Aim of study	To correlate the abundance of a suite of moorland breeding birds species with measures of vegetation composition and structure and to assess the implications in terms of the likely impacts of vegetation change on moorland bird populations
	Study design	2
	Quality score	+
	External validity	++
Population and setting	Source population	Source population is UK upland moorland and associated breeding bird assemblages. A brief review of previous, largely broad-scale, correlative studies given.
	Eligible population	Study area is moorland of S Scotland and N England. Representative of UK moorland breeding bird habitat
	Inclusion and exclusion	Sample of 85 large-scale plots selected from within National Countryside Monitoring Scheme sites selected at random but stratified by heather cover, across ten upland

	criteria	areas. Some additional sites from 1990 heather map from Landsat imagery. Due to access refusal the sample was made up by 32 non-random plots. Unintentional bias is discounted due to reasonable predictive power of resulting model.
	Setting	Various moorland areas across southern Scotland and N England.
Methods of allocation to intervention/control	Methods of allocation	N/A correlative study – no treatments imposed
	Intervention description	Non-experimental. Variables reflect background management including grazing.
	Control/comparison description	N/A
	Sample sizes	85 2km x 2 km squares
	Baseline comparisons	N/A
	Study sufficiently powered	The possibility of intercorrelation and the possibility of type I error (detection of spurious effects) is discussed.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Breeding bird density
	Secondary outcome measures	Key variables influencing breeding bird occurence
	Follow-up periods	N/A – Survey approach with three observations in one breeding season
	Methods of analysis	Modelling approach building minimum adequate models (MAMs) relating bird abundance to non-veg variables by stepwise selection. Veg variables then inserted to

		determine additional significant effects. The siginificant variables then incorporated into existing models and predictive power tested against one of the ten hill areas. The approach attempts to reduce the problems of intercorrelation – the possibility of type I errors is seen as less serious than possible type II errors (failure to detect some vegetation effects) through intercorrelation.
Results		Analysis was performed on nine species that were recorded in at least 30 plots. Stage 1 models (non-vegetation variables) explained 24-74% of the deviance in bird abundance. For some species, particularly red grouse and curlew, the effect of non-vegetation variables left little variation to be explained by adding vegetation variables. For snipe, meadow pipit and wheatear, non-vegetation variables accounted for very little variation. There was considerable variation in the effects of vegetation composition and structure after accounting for non-vegetation variables. For example none of the vegetation variables was significantly correlated with residual wheatear abundance despite the non-veg variables explaining just 24% of variance in abundance whilst skylark was correlated with 11 veg variables. Residual red grouse abundance was most highly correlated with variation in dwarf shrub cover, and negatively correlated with short grass cover. Residual golden plover abundance was positively correlated with short dwarf shrub cover and negatively with vegetation density. Curlew and snipe tended to be associated with structural heterogeneity and regression coefficients indicated that the latter was more abundant where tall vegetation cover was approximately 50%. All three waders were associated with plants indicative of wet conditions. Skylark abundance indicated an avoidance of heath in favour of grassland and meadow pipit tended to be more abundant on grass-heath mosaics. Whinchat and stonechat tended to be associated with tall vegetation with a preference for bracken and heather respectively. Quadratic relationships for both species with cover of fine leaved grasses indicated a preference for 20% cover. From these findings loss of heather is likely to reduce habitat availability for red grouse

		and stonechat of the nine species studied. Change in vegetation structure and overall heterogeneity may have greater direct effects on several of the species studied than changes from heather to graminoid dominance. Species which require areas of short, open vegetation or structural heterogeneity, notably the waders, may be adversely affected by further declines in livestock numbers
Notes	Limitations identified by author	Limitations of modelling approach and possible intercorrelation effects. Certain relationships identified may represent surrogates for effects of other correlated variables.
	Limitations identified by review team	No data included on grazing pressure e.g. annual average rates for sites. Cover measurements made along a 1m cane rather than quadrat. May not be reliable.
	Evidence gaps and/pr recommendations for further research	Research to verify relationships between birds and vegetation variables identified and assess their applicability to other areas of the UK
	Sources of funding	RSPB

Quality Assessment Checklist: Quantitative Study Observational / Correlation v2.0

Name of Evidence Review: ___Upland_____

Name of Review Sub-topic (if any): ___Moorland grazing_____

Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?
Study Citation	Pearce-Higgins, J. W. & grant, M. C. (2006). Relationships between bird abundance and the composition and structure of moorland vegetation. Bird Study, 53, 112-125
Study Design Category	2
Assessed by & when	D Martin 3/1/13

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐+	Comments: Source population is UK upland moorland and associated breeding bird assemblages. A brief review of previous, largely broad-scale, correlative studies given.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Comments: Study area is moorland of S Scotland and N England. Representative of UK moorland breeding bird habitat
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++	Comments: Sample of 85 large-scale plots selected from within National Countryside Monitoring Scheme sites selected at random but stratified by heather cover, across ten upland areas. Some additional sites from 1990 heather map from Landsat imagery. Due to access refusal the sample was made up by 32 non-random plots. Unintentional bias is discounted due to reasonable predictive power of resulting model.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐NA	Comments:
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++	Comments: Direct measurement of vegetation variables included cover of species or groups, height of different components of vegetation, height variability and vegetation density. Graminoid tussock index also recorded. All measurements sampled systematically on transects at two periods. Other environmental variables included peat depth and presence of burning. Predator (crow) abundance estimated and keeper density on estates covered. No estimates of grazing pressure included.
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NA	Comments:
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐++	Comments: Large sample likely to help avoid bias. Known confounding factors such as proximity to forestry
2.5 Is the setting applicable to the UK?	☐++	Comments:

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐++	Comments: Bird surveys following established methods. Three visits per plot, with min of seven days between visits. Entire plot covered to within 100m. Skylark and meadow pipit from two 1km line transects. Not undertaken on wet or windy weather. Assessments of observer variation made and found significant for Curlew, so allowed for in analysis. Red grouse counts validated against counts with dogs on a sub-sample.
3.2 Were all outcome measurements complete?	☐+	Comments: MP and skylark counts missed on 9 plots due to poor weather.

Were all/most of the study population that met the defined study outcome definitions likely to have been identified?		
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++	Comments:
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++	Comments:
3.5 Were there similar follow up times in exposure and comparison groups?	☐NA	Comments:
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐NA	Comments: Correlative study based on assessment in one season

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	Comments: The possibility of intercorrelation and the possibility of type I error (detection of spurious effects) is discussed.
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐++	Comments: A total of 45 vegetation variables (including 20 plant taxa) and 38 environmental or management variables. Collinearity tested by initial correlations between bird abundance and closely related variables, and selecting the most significant of the group for inclusion in multivariate analysis. This reduced to 23 veg and 15 non-veg variables
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐++	Comments: Modelling approach building minimum adequate models (MAMs) relating bird abundance to non-veg variables by stepwise selection. Veg variables then inserted to determine additional significant effects. The siginificant variables then incorporated into existing models and predictive power tested against one of the ten hill areas. The approach attempts to reduce the problems of intercorrelation – the possibility of type I errors is seen as less serious than possible type II errors (failure

		to detect some vegetation effects) through intercorrelation.
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	Comments: All significant relationships presented at p<0.05. The limitations of the approach and the possibility of type II error is discussed in the paper.
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐+	Comments: Large sample, well designed study but the large amount of multiple variables opens possibilities of autocorrelation. This is recognised and adjusted for as far as possible.
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐++	Comments: Large sample of moorland, with widespread moorland bird species.

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	What are the effects of grazing regimes and stocking rates on the maintenance and or restoration of moorland biodiversity and ecosystem service delivery?

Study details	Authors	Pearce-Higgins, J. W. & Grant, M. C.
	Year	2002
	Aim of study	To examine the effects of grazing on skylark and meadow pipit density at a range of scales, through examining the relationships with a number of habitat variables.
	Study design	Transect survey within partially stratified random squares
	Quality score	-
	External validity	+
Population and setting	Source population	Unenclosed moorland habitat in S Scotland and N England
	Eligible population	2km^2 sample areas across the source population
	Inclusion and exclusion criteria	None specified
	Setting	S Scotland and N England moorland. No more specific geographic limits given
Methods of allocation	Methods of allocation	Survey squares allocated on a partially random basis, based on heather cover

to intervention/control	Intervention description	Survey approach – no intervention applied, but survey will cover a range of current and historic grazing pressures.
	Control/comparison description	No control – survey covers a range of conditions
	Sample sizes	Meadow pipit and skylark recorded in 76 of 85 sample areas. All occurrences of the species along 2x1km transect recorded. 20 vegetation measurements per transect and additional 80 points over sample square
	Baseline comparisons	One-off study, not a baseline-resurvey approach
	Study sufficiently powered	No power analysis given, but large sample. Power to detect significant effect will vary between variables. Possibility of type II error through under-occupancy of suitable habitat.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Outcomes are bird densities and relationship to vegetation and environmental variables at three scales. Assumptions made in calculating breeding density – halving number of birds from 1st visit.
	Secondary outcome measures	None
	Follow-up periods	Not really relevant. The study is survey based, measuring the impacts of current and historic grazing, rather than a treatment applied for a set period.
	Methods of analysis	Modelling approach – probably appropriate for range of variables measured. Spatial autocorrelation in bird distribution tested for. Analysis at transect scale used number of birds as dependant variable, with vegetation measures as independent variables.
Results		Estimated densities of skylark and meadow pipit higher than from other upland studies. Skylark densities highest on short grass moorland, and negatively associated with

		bracken and high cover of Molinia. Meadow pipit abundance peaked at intermediate heather cover at plot and transect scales. Pipits were associates with cover of tall rushes and wavy hair grass, but negatively correlated with heath rush. Suggestion that grazing may have had a role in creating conditions for skylark, and meadow pipit where heather has become fragmented (holds true for mean and breeding density for both species). Pipits appear to favour tussocky vegetation, in a fairly uniform sward, at the sample point scale. At the transect scale maximum densities are associated with 30% heather cover. Since pipits are a favoured prey of Hen Harrier, the suggestion is made that managing for meadow pipits can benefit this species.
Notes	Limitations identified by author	Density estimates differ from other national studies – may be down to survey methodology.
	Limitations identified by review team	Links between some of the surrogate vegetation measures and actual grazing pressure may be weak, and/ or influenced by other environmental factors.
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	RSPB, with support from SNH and GCT

Quality Assessment Checklist: Quantitative Study Observational / Correlation v2.0

Name of Evidence Review: _____Uplands_____

Name of Review Sub-topic (if any): _____Moorland Grazing_____

Review Question	
Study Citation	Pearce-Higgins, J. W. & Grant, M. C. (2002). The effects of grazing-related variation in the habitat on the distribution of moorland skylarks *Aulauda arvensis* and meadow pipit *Anthus pratensis*. Aspects of Applied Biology 67
Study Design Category	
Assessed by & when	D Martin 12/10/12

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐-	Comments: Not described in any detail – taken to be the range of unenclosed moorland habitats of S Scotland and N England.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+ ☐	Comments: 85 plots selected using a partially stratified random sample based on heather cover. Approach not fully described in this paper, but refers to a paper 'in prep'.
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐+	Comments: Each 2km^2 plot sampled along two 1km transects 600m apart. Unsure how transects were selected. All birds of the target species were recorded, and vegetation measurements at 50m intervals.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐+	Comments: Correlative survey approach rather than experimental. Exposure is the range of prevailing environmental and grazing conditions. Bias minimised through large sample, full survey of bird presence. Vegetation measured systematically, at 50m intervals. Could introduce unintended bias?
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐-	Comments: Various height measures and indices of biomass and structural heterogeneity derived. The various habitat condition measures are presented as surrogates of historical grazing pressure. However, also reflect current grazing, and underlying environmental and soil conditions.
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NR	
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐-	Comments: May be other management factors related to sporting management on heather dominated or heather mosaic moors – e.g. predator control
2.5 Is the setting applicable to the UK?	☐++	Comments: Yes – all sites in Scotland or N England

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐-	Comments: Outcomes are bird densities. Assumptions made in calculating breeding density – halving number of birds from 1st visit. There is an indication of significant observer effect at the transect scale in the model (table 4)
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐++	Comments: Yes

3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐+	Comments: Main outcomes are density and distribution of the two bird species, in relation to different habitat types.
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐+	Comments: Yes, although assumptions made about how observed density relates to breeding density.
3.5 Were there similar follow up times in exposure and comparison groups?	☐-	Comments: The study is survey based, measuring the impacts of current and historic grazing. In this respect the sites have been subject to a range of different exposures, and the study attempts to measure vegetation parameters as grazing surrogates, that have resulted from different (unquatified) grazing levels and patterns over different timescales
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐+	Comments:

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	Comments: No power analysis, but a reasonably large sample size. Modelling approach used. And significance of relationships presented. A chance of type II error through under-occupancy of suitable habitat.
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐+	Comments: Yes, a range of explanatory variables considered. However basis of some surrogate measures quite theoretical, e.g. tussock index, vegetation biomass.
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐++	Comments: Modelling approach – probably appropriate for range of variables measured. Spatial autocorrelation in bird distribution tested for.
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful?	☐++	Comments: Significance of relationships in model given to up to 4 decimal places

Were confidence intervals and or p-values for the effect estimates given or calculable?		
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐-	Comments: Bias minimised through large sample, full survey of birds on transect, systematic vegetation sample etc. However the links between many of the surrogate measures to current and historic grazing not adequately explained or tested. Undoubtedly habitat preferences identified, but link to grazing relies on various assumptions.
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐+	Comments: Largely, due to large sample, but source population and how representative the sample is not clear.

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services? b) What methods of stocking rate calculation, or setting grazing regimes, consistently provide regimes that maintain or restore moorland biodiversity, and what are the key parameters that calculations should include?

Study details	Authors	Poulton
	Year	2011
	Aim of study	To analyse data from a number of grazing surveys, and evaluate what changes have taken place over time, following stock reductions.
	Study design	1
	Quality score	++
	External validity	++
Population and setting	Source population	Meta analysis of a large number of surveys on English moorland
	Eligible population	The surveys are biased towards heavily grazed sites as they are related to a policy of reducing overgrazing
	Inclusion and exclusion	sites reported as potentially overgrazed. Sites were identified by advisers, but a

	criteria	subsequent sample study suggested these sites were at the heavier grazed end of the spectrum
	Setting	A large number of overgrazed moorland sites in English uplands – NW, Yorkshire, S Pennines, Shropshire Hills and SW moors.
Methods of allocation to intervention/control	Methods of allocation	N/A
	Intervention description	All sites subject to stock restrictions based on
	Control/comparison description	N/A
	Sample sizes	141 sites, 247 site visits, 26 466 quadrats
	Baseline comparisons	Many surveys were carried out to identify overgrazing, and therefore from before stock reductions. Part of the analysis looks at change within sites with repeat surveys.
	Study sufficiently powered	Yes
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Grazing Index, proportion of heavily grazed features, frequency and percentage cover of species and groups, herbage height.
	Secondary outcome measures	
	Follow-up periods	Surveys over a 13 year period, variable time between initial survey and resurvey, where present.
	Methods of analysis	Within site analysis of change where there were more than one visit and a between site analysis. Within site, each variable was analysed using GLM, and second stage meta

		analysis of site results. This involved parametric t-testing of Beta coefficients from best fit regressions of variables at individual sites, against the null hypothesis that the mean is 0, i.e. no change. Corroborated using chi-squared test of trinary (-1, 0, +1) variables of regression slope from individual sites. Larger between-site analysis of site-year variables. Monte Carlo approach to GLM deriving two sets of beta coefficients from sub-sampling and randomised sub-sampling of years and used to test significance of change in the response variable. PCA of categorical data for dominant species.
Results		A number of surrogate management variables showed highly significant changes over time. Sheep and cattle/ pony dung declined as did heather grazing index (mean of 60% in 2000 to 40% in 2008) and proportion of heavily grazed features. Values of GI exceeding 70% were common in early surveys whilst only recorded on two site visits in the latter four years. Growth stage of heather declined (more younger heather) which may be related to burning. Moorland species and community variables do not so clearly reflect the changes. Cover and abundance showed very little change, the only significant change being a decrease in bare ground in the within-site analysis, although a marginally significant opposite effect was seen in the between-site analysis. Heather showed an increase in height in the within-site analysis but not between sites. Palatable grasses however seemed to decrease in height in the between site analysis.
Notes	Limitations identified by author	Inconsistencies in data, particularly recording null values rather than zeros. In most cases used 'as found' but obvious errors rectified. Design of data problematic –partial repeated measures with a large degree of non-independence, due to only some sites having repeat visits.
	Limitations identified by review team	

	Evidence gaps and/pr recommendations for further research	
	Sources of funding	Natural England

Quality Assessment Checklist: Quantitative Study Observational / Correlation v2.0

Name of Evidence Review: ____Uplands_____

Name of Review Sub-topic (if any): ___Moorland Grazing_____

Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services? b) What methods of stocking rate calculation, or setting grazing regimes, consistently provide regimes that maintain or restore moorland biodiversity, and what are the key parameters that calculations should include?
Study Citation	Poulton, S. (2011) Preliminary Analysis of grazing management data. Report to Natural England. BioEcoSS Ltd.
Study Design Category	1
Assessed by & when	D Martin 14/12/12

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐NA	Comments: Meta analysis of a large number of surveys on English moorland
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐- ☐NA	Comments: The surveys are biased towards heavily grazed sites as they are related to a policy of reducing overgrazing
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++	Comments: Same as eligible population – sites reported as potentially overgrazed. Sites were identified by advisers, but a subsequent sample study suggested these sites were at the heavier grazed end of the spectrum

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐NA	Comments:
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++	Comments: A range of variables identified from previous studies as being affected by grazing pressure. Mainly relates to dwarf shrub
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NA	Comments: Survey of impacts of prevailing grazing regimes
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐+	Comments: Large scale study/ sample helps account for climate, soil effects on dwarf shrub growth. 141 sites with a total of 247 site visits (surveys). Data from 26,466 quadrat locations
2.5 Is the setting applicable to the UK?	☐++	Comments: Range of sites from across English uplands

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐+	Comments: Objective measures of sward height, cover, utilisation etc from random quadrats, with a min of 80 on most sites, other than v small. Carried out by experienced surveyors.
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐++	Comments:
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++	Comments:

3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++	Comments:
3.5 Were there similar follow up times in exposure and comparison groups?	☐NR	Comments:
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐+	Comments: Studies over a 13 year period some but not all had repeat surveys, often after 2 years. Sometimes multiple visits. Likely to be long enough to detect change in some variables (structural) but not composition.

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐++ ☐+ ☐- ☐NR ☐NA	Comments:
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐++	Comments: Whole range of survey variables included in the analysis.
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐++	Comments: Within site analysis of change where there were more than one visit and a between site analysis. Within site, each variable was analysed using GLM, and second stage meta analysis of site results. This involved parametric t-testing of Beta coefficients from best fit regressions of variables at individual sites, against the null hypothesis that the mean is 0, i.e. no change. Corroborated using chi-squared test of trinary (-1, 0, +1) variables of regression slope from individual sites. Larger between-site analysis of site-year variables. Monte Carlo approach to GLM deriving two sets of beta coefficients from sub-sampling and randomised sub-sampling and used to test significance of change

		in the response variable. PCA of categorical data for dominant species.
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++	Comments: p values for t-tests and chi-squared. Mean and 95% CI of beta values for between site analysis.
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐++	Comments: The field studies are based on random sampling. Various analyses to take account of sites with repeat samples and change across all sites, and to account for different types of variables.
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐++ ☐+ ☐-	Comments: the findings on likely change in variables with similar magnitude of livestock reduction are generalisable to other moorland